Calhoun: The NPS Institutional Archive

Theses and Dissertations Thesis Collection

1985-09

Data Dictionary/Directory Systems, a tool in systems development life cycle

Drakoulis, Christos

http://hdl.handle.net/10945/21239

Calhoun is a project of the Dudley Knox Library at NPS, furthering the precepts and goals of open government and government transparency. All information contained herein has been approved for release by the NPS Public Affairs Officer.

Dudley Knox Library / Naval Postgraduate School
411 Dyer Road / 1 University Circle
Monterey, California USA 93943

http://www.nps.edu/library

DUDLEY KNOX LIBRARY
NAVAL POSTGRADUATE SCHOOL
MONTEREY, CALIFORNIA 93943

NAVAL POSTGRADUATE SCHOOL
Monterey, California

THESIS

DATA DICTIONARY/DIRECTORY SYSTEMS, A TOOL
IN SYSTEMS DEVELOPMENT LIFE CYCLE

by

Christos Drakoulis

September 1985

Thesis Advisor: N. R. Lyons

Approved for public release; distribution is unlimited

T226288

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)
Data Dictionary/Directory Systems, a Tool in Systems Development Life Cycle

5. TYPE OF REPORT & PERIOD COVERED
Master's Thesis
September 1985

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)
Christos Drakoulis

8. CONTRACT OR GRANT NUMBER(s)

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Naval Postgraduate School
Monterey, CA 93943-5100

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS

11. CONTROLLING OFFICE NAME AND ADDRESS
Naval Postgraduate School
Monterey, CA 93943-5100

12. REPORT DATE
September 1985

13. NUMBER OF PAGES
100

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)

15. SECURITY CLASS. (of this report)
UNCLASSIFIED

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution is unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

DD/DS: Data Dictionary/Directory System
IRM: Information Resource Management
SDLC: System Development Life Cycle

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)
Although data resources have some characteristics with other corporate resources, such as people, goods or money, they do not have two specific characteristics. They are not relatively scarce, neither are they inherently allocatable. However, they do have value, both positive or negative. The value is derived from the fact that the entire enterprise depends on their availability for the proper management of all the other resources. Organizations are now beginning to treat data as (Continued)

ABSTRACT (Continued)

a resource, which requires the same degree of administration and control as is involved in the management of other resources. A key component in this management and control is the Data Dictionary/Directory System. It is a useful tool, throughout a System Development Life Cycle.

Approved for public release, distribution unlimited

Data Dictionary/Directory Systems, a Tool in Systems Development Life Cycle

by

Christos Drakoulis
Lieutenant Commander, Hellenic Navy
B. S., Hellenic Naval Academy, 1970

Submitted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE IN INFORMATION SYSTEMS

from the

NAVAL POSTGRADUATE SCHOOL
September 1985

ABSTRACT

Although data resources have some common characteristics with other corporate resources, such as people, goods or money, they do not have two specific characteristics. They are not relatively scarce, neither are they inherently allocatable. However, they do have value, both positive or negative. The value is derived from the fact that the entire enterprise depends on their availability for the proper management of all the other resources. Organizations are now beginning to treat data as a resource, which requires the same degree of administration and control as is involved in the management of other resources. A key component in this management and control is the Data Dictionary/Directory System. It is a useful tool, throughout a System Development Life Cycle.

TABLE OF CONTENTS

LIST OF FIGURES

I. INTRODUCTION

In the early 1950s the first general purpose electronic digital computers were introduced. They promised to simplify, ease and enhance life in organizations of all kinds. Since then, they kept their promises. Their use in all types of organizations and commercial enterprises has grown at an astounding rate. As a result, it is hard to imagine how these enterprises could function today without computers.

The "Computer Revolution" during the last few years has had a tremendous impact on the data and information used by enterprises. It resulted in a corresponding "Information Revolution". Using the computer, an enterprise can store and retrieve vast amounts of data and information. Being able to process data and use information effectively is vital to both business and government organizations. The repository of data and information is called a "Data Base". It is the foundation of the entire computer-based processing system for an enterprise. Data and information are very important to enterprises. Thus, any improvement in the way data are handled is going to improve the ability to manage an enterprise and the quality of services it provides.

Database technology has emerged since 1960s in the data and information management field. Data integration and data independence were the focal points of this technology. This orientation of data management increased the productivity of the systems, particularly by improving data accessibility and reducing data redundancy. Additional thinking about organized approaches for the creation and support of data and information, emerged a new concept. The concept of data as an important corporate resource. Thus, data and information resource management have become today the current trend and focal points of data processing.

Data Dictionary/Directory Systems (DD/DS) made their debut in the mid-1970s with the management of database definitions as a central goal. They protected the integrity of definitions and also accurately documented the data resources. What started as a tool for description and documentation of data within a database, soon became a service for data resource control. Thus, the DD/DS became a primary administration tool, supporting in a comprehensive manner, the logical centralization of data resources. Both in theory and in practice, a DD/DS in its implemented form can support a manual system as well as a highly sophisticated automated one. It may also exist as an independent system as well as a depended one on a specific database management system.

This thesis will explore and describe the role of the DD/DS in the systems development life cycle (SDLC). To better comprehend this role, key concepts of today's complex data processing environment will be discussed. Included in this discussion is data and information concepts since their early stages, database concept, system development life cycle concepts, and information resource management concept. A key component in the SDLC is the DD/DS. Its usefulness and in particular the benefits that can be derived from its use are explored and emphasized. The final discussion will be on a survey of the currently available DD/DS packages, offered by vendors, and their problems.

II. PRINCIPLES AND CONCEPTS

A. INFORMATION

A foundation for a modern enterprise is the information needed for its survival and prosperity. It is a resource parallel in importance to labor, land and capital. Information may be defined as:

> data that have been processed into a form that it is meaningful to the recipient or user and is of real or perceived value in current or prospective decision processes. [Ref. 1]

Information is a term which must be distinguished from the term data. Data are facts, records of an event that has occured or is about to take place. They undergo a transformation involving infusion with intelligence and they become information. Thus, although information originates from data it differs from them because all data may not become information. Also, what is information for one person may not be for another. Information adds to relevant knowledge, reduces uncertainty, and supports the decision-making process in an organization.

Senn [Ref. 2] further describes the general attributes of information, both for an item of information and for a set of information (Fig. 1).

The information used in managing an enterprise, whether applied in a communication sense or in a decision-making context primarily comes from two sources: Internal and

Attributes of an Item of Information	
Accuracy	Information is true or false accurate or inaccurate.
Form	It is the actual structure of information. It includes the dimensions of quantifiability, level of aggregation and medium of presentation. Often a selection of one or the other alternate forms is dictated by the situation.
Frequency	Is a measure of how often information is collected, needed or produced.
Breadth	It defines its scope. Some information may be broad in scope, covering a large area of interest. Other information may be narrow in its scope. Usage determines the necessary breadth.
Origin	It is the source from which it is received, gathered, or produced.
Time Horizon	Information may be oriented toward the past, toward current events, or toward future activities and events.
Attributes of a Set of Information	
Relevance	Information is relevant if it is needed for a particular situation. Information needed at one time may not be relevant now. Likewise information obtained "just in case it is needed" is not relevant.
Completeness	Complete information provides the user with all that needed to know about a particular situation.
Timeliness	Timely information is information that is available when it is needed. Further, it has not become out-dated through delay.

Figure. 1--The Attributes of Information [Ref. 2] external. Identifying these sources and being able to evaluate the reliability of each is an important task for a manager. The manner with which this task is accomplished affects the value of the information and the purpose for which it will be used.

Information processing as decision-oriented data processing is aimed at transforming data into information needed by managers. Today this transformation commonly accomplished via computers, which with computer programs, carry out the calculations, manipulations and data reduction.

B. INFORMATION SYSTEMS

Information used in managing an enterprise can be categorized as belonging to three levels of hierarchy [Ref. 3] as illustrated in Fig. 2.

On the first level there is the repetitive, predictable, routine, frequently produced, and frequently accessed information. This information is used for day-to-day operation by the first-level management, such as the accounting, payroll and credit departments and is known as operational information. This basic needed information is the most amenable to automation in an enterprise.

On the second level there is the functional information. At this level similar information types are grouped together into functional units. This provides operating management with a degree of control and a broader scope of information about the business operation.

At the third level there is the executive-level information. It is designed to provide management with information supporting enterprise-wide policy-making activities and high level planning. This information is characterized by a high

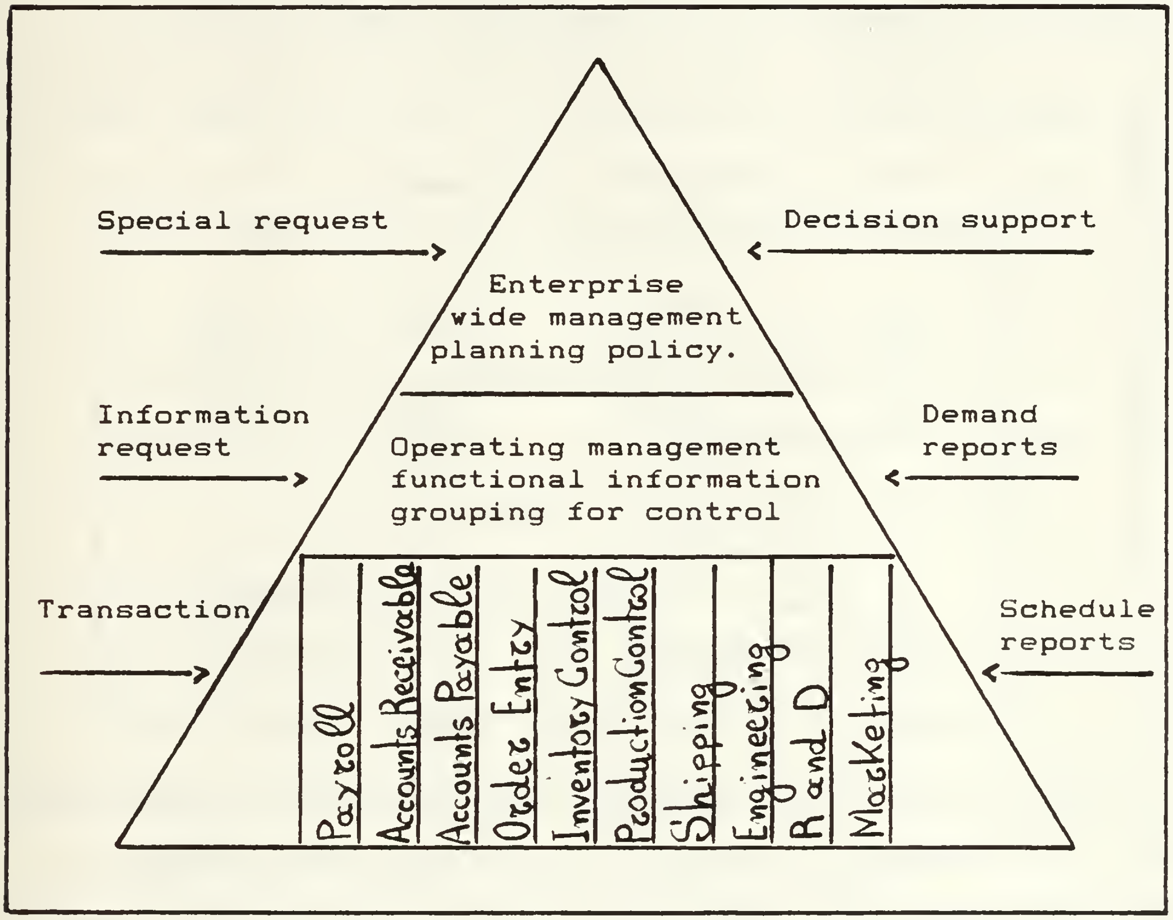

Figure 2.--The Information System Hierarchy [Ref. 3]

degree of summarization. Automation at this level enables sophisticated manipulation of available information, higher degrees of integration, and more complex analytical reports on a more timely basis.

Due to the needs for information on a continuing basis, it is necessary to develop a subsystem for processing and handling the information resource alone. An Information System may be defined as:

> a system which uses personnel, operating procedures, and data processing subsystems to collect and process data and disseminate information in an organization [Ref. 4].

Hardware, software, data, personnel and procedures are the basic elements of an information system (Fig. 3).

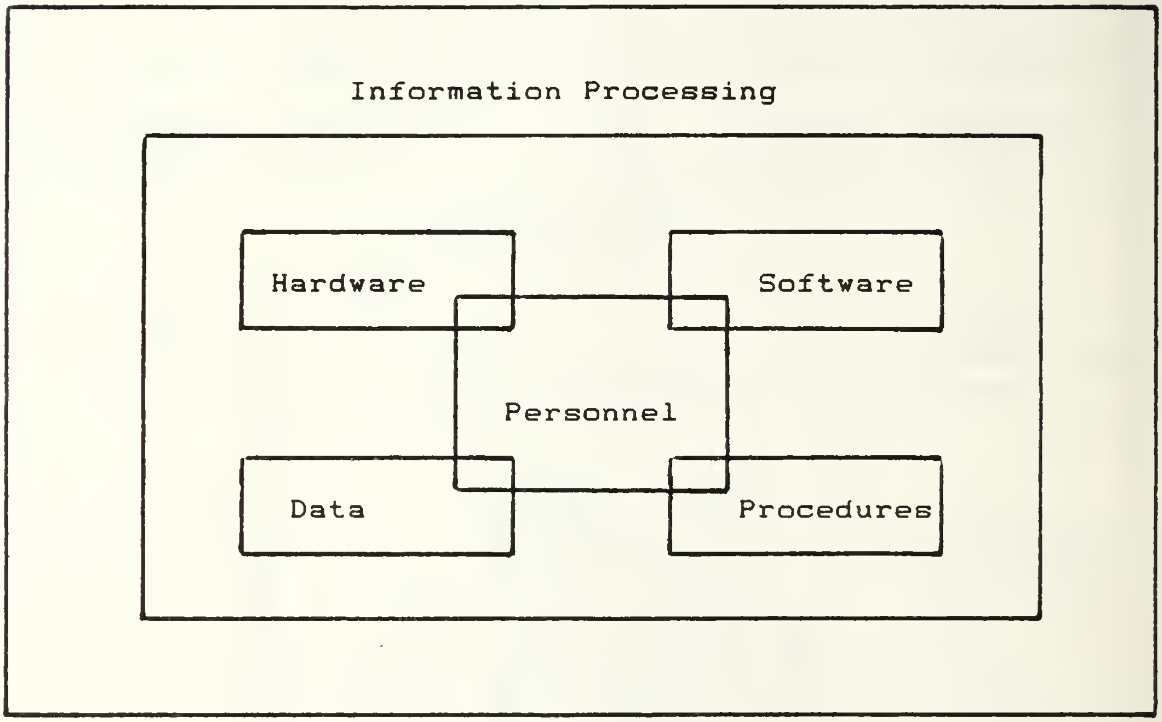

Figure 3--Elements of an Information System

Information systems have evolved from manual systems to automated systems; database systems; on-line systems; to user-friendly, and highly sophisticated systems of today's. Each stage of evolution involved a transition brought about a changing environment in response to problems. Each stage solved some problems and reduced personnel requirements, but opened the door to a whole new set of problems. Each stage shortened or broadened the information cycle, but opened up possibilities and opportunities not feasible previously. In a few years when it is expected that the fifth generation computer systems will be in wide use, information processing systems will be the central tools in all areas of social activity. They will include economics, industry, art and science, administration, international relations, education, culture and daily life. Such information systems will be required to meet the new needs generated by environmental changes. They will not only be expected to play active roles in the resolving of anticipated social bottlenecks but also to advance society along a more desirable path through the effective utilization of their advanced capabilities.

C. FILE ENVIRONMENT

The success of an information system depends on proper management of data. Even if the hardware and software elements are well designed and operating properly, the value of the system will be marginal if the underlying data are not reliable. Similarly, if the application programs that need the data cannot get at them, the system may be limited in usefulness.

When organizations first introduced computers, they developed the so called file systems. A file system whether on cards, tapes or disks may be defined as:

> designed to receive, store and produce specific information/output that requires specific input. A file

system is a collection of individual files of organized records that serves the current information needs of specific application [Ref. 5].

In a file system the smallest element in the hierarchy of data is the data item. The data item is a fact or statement about an entity (person, place, thing, or event). For the data items to be distinguished from each other, should have clearly identified characteristics of name, size and type specification. Related data items are then grouped together to form records, which in turn are grouped into files. Data records may be fixed or variable, whether the number and position of data items in the record are fixed or variable in length. When finally the file is established and its definition is decided, it becomes fixed and all records must conform to the definition. Data items and records that form a file in a student situation, are shown in Fig. 4-7.

In the file oriented environment two types of files were commonly developed and used. The first, Master files, hold historical overview of past events, and also show the status of items of current interest. The second, Transaction files, capture data about occuring events, and are used to update the Master files. Into these files the data are stored sequentially or randomly, and the relation between items and records is physical.

As organizations grow and gain experience with data processing they discovered the major drawbacks of the file environment. They had been engaged to use files for only one

Entity : Student

Data Item Name	Size	Type	Value
Student Name	16	X(alphanumeric)	Tim Walker..
Identification Number	5	9(numeric)	

Figure 4.--Relationship between an Entity and two Data Items

Record : Student (a 60-character record)

Data Item Name	Size	Type	Value
Student Name	16	X	Tim Walker.......
Identification Number	5	9	22444
Address	20	X	311 Omaha St.
City	12	X	Monterey
State	2	X	CA
Zip Code	5	9	93940

Figure 5.--An Instance of the Record Student

Record	Student Name	ID No	Address		City	State	Zip Code
1	Tim Walker	22444	311	Omaha St.	Monterey	CA	93940
2	John Nolan	28533	121	River St.	Monterey	CA	93940
3	Bill Perry	33555	132	Lake St.	Monterey	CA	93940
4	Nancy Norm	12333	44	Demon St.	Monterey	CA	93940
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.

Figure 6.--A Student File (collection of records)

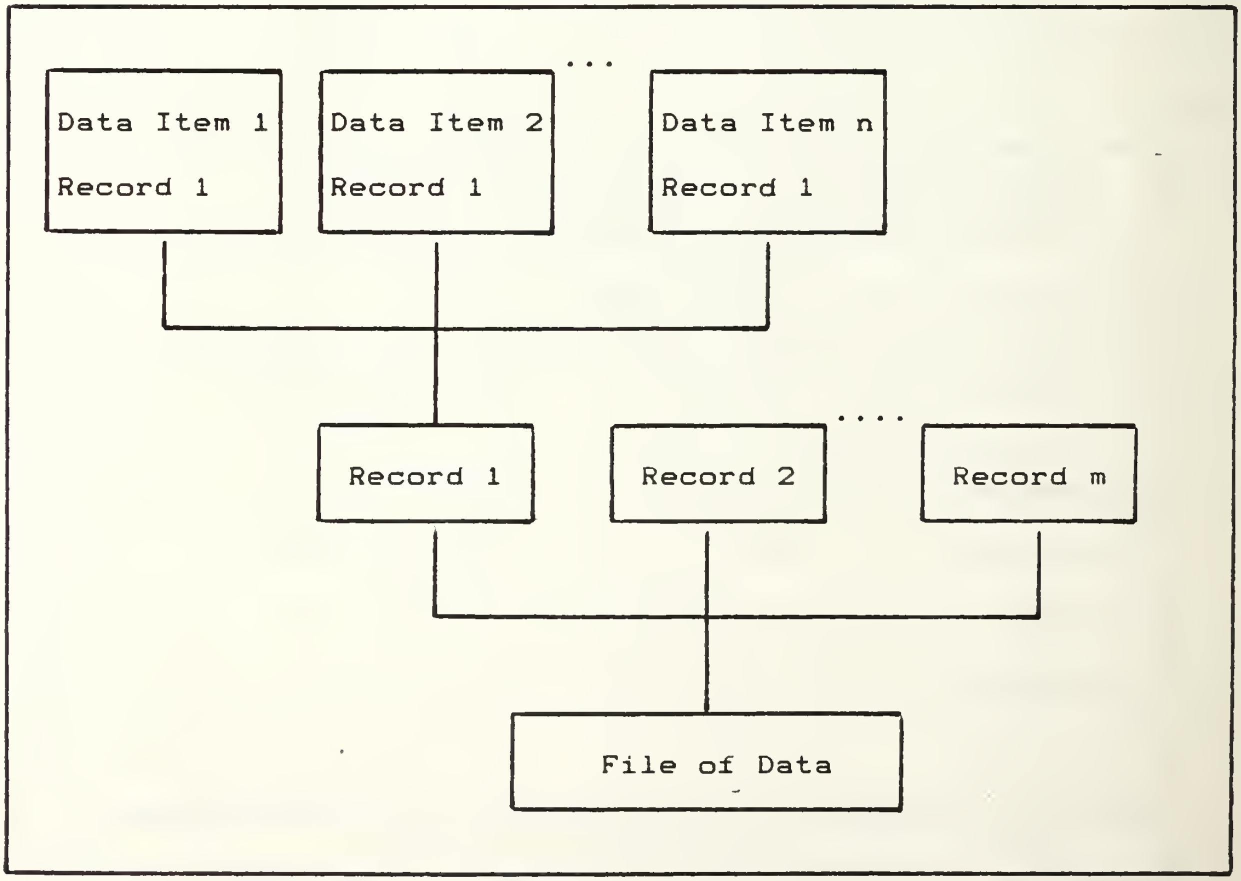

Figure 7.--Data Hierarchy

or two applications, and they realized there was a limited amount of data sharing across applications. Thus, they were forced to have data and file redundancy. Later, data and file duplication led them to problems of data integrity and most of the times to problems of data inconsistency and inaccessability. Meanwhile, the cost of labor was increasing steadily, while the cost of computers was decreasing dramatically. Now, there was the potential for trading people resources for machine resources. For these reasons, and also to solve the kinds of problems that the file environment created, database systems were developed.

D. DATABASE ENVIRONMENT

The basic difference between a conventional file or application system and a database system is in the philosophy. A database is a collection of related data about an enterprise that can be processed as an integrated whole. The database does not store information as information, but it stores data to be used in generating multiple levels and types of information. In a database the data definitions and the relationships between data are distinguished from the procedural statements of a program. The database is a self-describing collection of integrated files [Ref. 6], because it contains within itself, a description of its structure. To understand and appreciate these concepts, consider the systems shown in Fig. 8. In the file processing system each

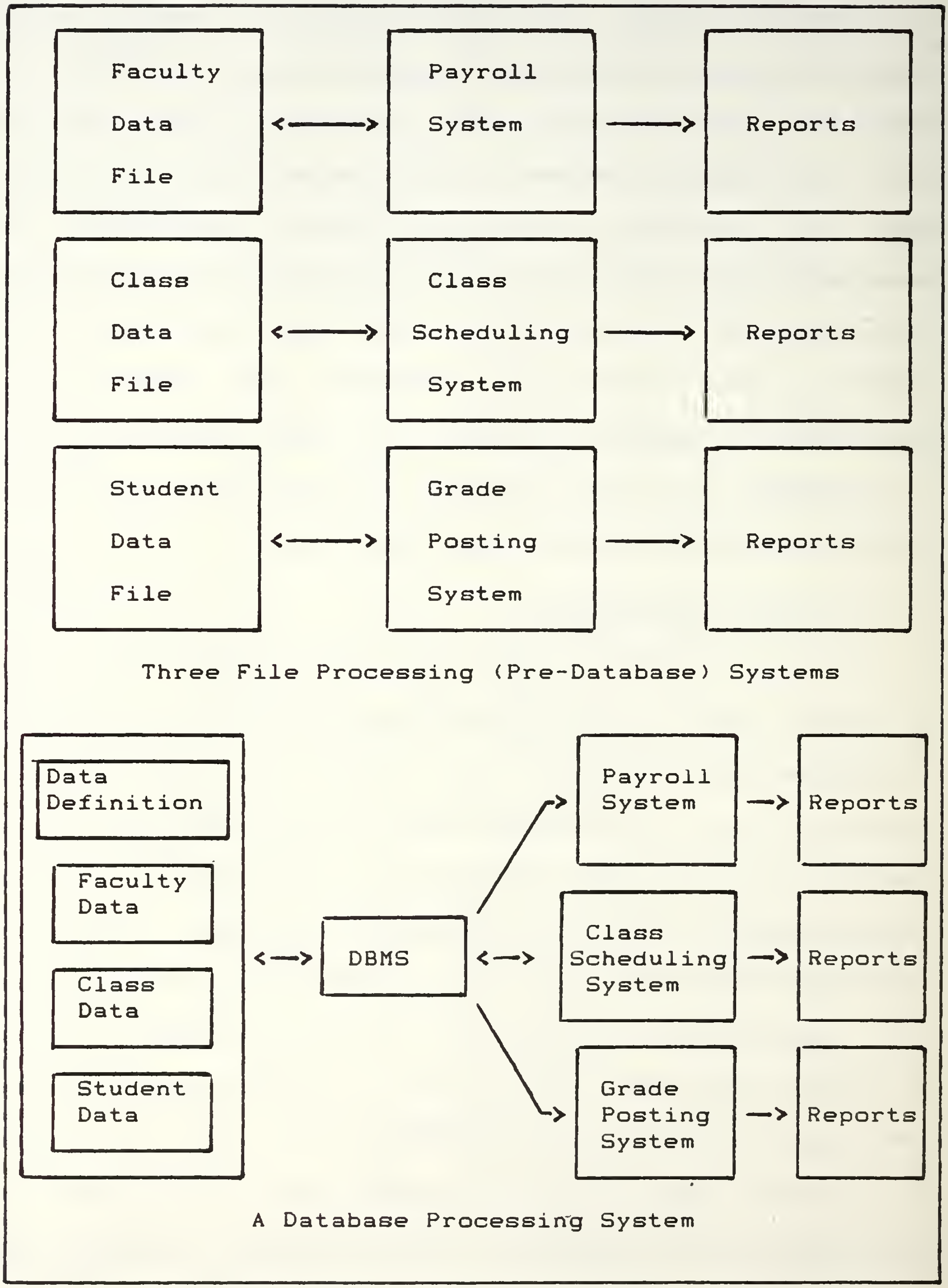

Figure 8.--Difference between a File and a Database System

file exists independently. To obtain combined information, a new program must be written extracting data from corresponding files. Instead, in a database system the files have been integrated into a database, that is, processed independently from the application programs. An intermediate system is also used called database management system (DBMS) which is a large and complex program. The DBMS stores data and also a description of the format of data and is called upon by the programs to access the database.

The database approach offers a lot of advantages. These advantages and the corresponding disadvantages of database processing can be seen in Fig. 9. Nevertheless, even with these considerable disadvantages, the advantages of using database technology make it extremely desirable in today's data processing environment.

An organization database system has five components: (1) Hardware, (2) application and utility programs, (3) data, (4) people, and (5) procedures. People for a database system are considered the users, operations personnel, systems development personnel and the database administration staff (DBA). DBA serves primarily as a protector of the database, resolving user's conflicts.

To design a database, one requires conceptual representations of the real world. The database design can be characterized as a scientific, intuitive, artistic, and iterative process. It is divided into two phases: logical

Advantages of Database Processing

- More information can be produced from the same amount of data
- Consistent information can be supplied for the decision making process
- Elimination or reduction of data duplication
- Program/data independence
- Application programs can be developed, maintained, and enhanced faster and more economically with fewer skilled personnel
- Better data management
- Security controls can be applied
- Representation of record relationships

Disadvantages of Database Processing

- Expensive, due to DBMS more hardware and higher operating costs
- Complex
- Backup and recovery difficult
- Integration of data, and hence centralization, increases vulnerability to failure

Figure 9.--Advantages and Disadvantages of Database Processing

design and physical design. A logical database design specifies the logical format of the database. The records to be maintained, their contents, and relationships among those records are specified. The logical design is usually called conceptual schema, or logical schema [Ref. 7]. The physical design is the phase of transforming the logical schema into the particular data constructs that are available with the DBMS to be used.

Database design should satisfy today's anticipated as well as future needs for information. It should be easy to modify, and expandable. Before inserting any data in the database, the data should be examined for validity and remain correct once is stored. Finally, it should provide security and privacy facilities to different users. Only authorized users should have access to the data stored in the database.

A database provides three views of data: (1) schema or (conceptual view), (2) subschema or (external/user) view, and (3) internal (physical view). The schema is the complete logical view of the data and describes all the data in the database. The subschema defines a subset of the schema which is accessed by a specific application program or user. Subschemas can overlap each other and can also reorganize the schema, depending on the DBMS used. The physical view is the form of the data as it is arranged and allocated to files. All these views must be defined before the database processing occur. Usually the DBA defines the schema and subschema. When the database is defined , the physical view is created automatically by the DBMS.

Having multiple views of data means that even though the data is centralized and shared, it can be tailored to the needs of the application. Then data can appear to each user in a more familiar and useful format.

E. SYSTEMS DEVELOPMENT LIFE CYCLE

When information systems are to be installed into organizations, they must be the product of careful planning. Systems developers must have in mind, that they have to provide the right system and also that the system must work right. From the time the need for an information system is first perceived to the time that the system is actually delivered, there are many stages that take place. These stages comprise the Information System Development Life Cycle (SDLC) and commonly are:

(1). Initiation phase or system planning phase.
This stage deals with formulation of a master plan of the system, (i.e. perception of the need, clarification of purpose, technical, economical, and operational feasibility).

(2). Requirements definition and analysis phase.
The system is first partitioned into subsystems in order each part can be studied effectively. Then user information needs are determined and described, and detailed system requirements are established.

(3). Logical system design phase.
The functional specifications of the system are formulated, (specification of procedures, input/ output, files and databases).

(4). Physical system implementation phase.

During this stage programs are coded and constructed. Record formats, data structures, files and databases are developed, and specific hardware devices are selected.

(5). Testing phase.
The implementation of programs, procedures, must be tested to ensure that the system is running properly.

(6). Implementation and evaluation phase.
During this stage users are trained, and evaluations are made of the system and of how it operates.

(7). Maintenance phase.
The SDLC is not completed after the installation and implementation phase. Improvements must be made continually to correct errors, in order to meet new management needs, or to take advantage of new technological improvements.

The phases of the SDLC are fairly universal and are accepted by management and the data processing community in general. However, in this high level categorization, there is no clear beginning and end for each phase.

No system can be effective if it does not meet management and, especially, user needs. Guidelines that analysts and designers must follow to avoid pitfalls and problems during the SDLC can be summarized as:

(1). The development of an information system should be tied to overall organization goals and objectives, whether these are profit maximization, cost minimization, or growth.

(2). Useful approaches for system development are the top-down and bottom-up approaches or a combination of both. The top-down approach is more effective to generate the general scope on how the system will evolve to support organization goals and objectives. The bottom-up approach may be followed within the overall development process.

(3). Before a systems request is approved and included in the master plan, assurance is needed that it can be accomplished within reasonable technical, economical and operational constraints.

(4). The determination and specification of user requirements must be accurate and complete, otherwise the new system has a great potential of failure.

(5). The system should be designed in the most cost effective and efficient manner. The design should provide efficiency, accuracy, and flexibility. The logical design requires a knowledge of the hardware and equipment that will be available for the project as well as, insight into potential users and purposes of the output. Users and systems developers alike should understand how the system was created and the techniques used to design it.

(6). Software development which is time-consuming, costly and an error-prone process must be iterative with feed-back from each stage to the previous one.

(7). Evaluating the new system is a critical step in learning how the system is operating and where changes must be made. Evaluating the system impacts involves looking at performance and at the effect of the applications within the information system.

(8). Finally the operation and maintenance stage of the SDLC counts for the major portion of the total cost. It is also an iterative process where system persons must cycle back and forth, acquiring additional information about design questions as the needs arise or as the problems change.

F. INFORMATION RESOURCE MANAGEMENT

1. Definition

Modern organizations are becoming increasingly complex, because of size growth, increased specialization, higher technology levels in products and processes, and the changing structure due to internal and external pressures. The proliferation of computers, and the introduction of automated information systems into management has been a major development providing a new tool for improving operations and planning. On the other hand, it has created an increasing need to manage these systems more effectively and efficiently. Resources into an organization are managed to optimize their utilization. To manage resources means to plan for, allocate, maintain and conserve, prudently exploit effectively employ, and integrate those resources. But primarily it is necessary to understand the resource, to know when and what is available, its source and also its destination. The Information Resource Management concept (IRM), was the result from the recognition for the need of managing information like any other resource in the organization. IRM may be defined as :

> whatever policy, action/procedure concerning information (both automated and non automated), which management establishes that serve the overall current and future needs of the enterprise. Such policies, would include consideration of availability, timeliness, accuracy, integrity, privacy, security, auditability, ownership, use and cost-effectiveness [Ref. 8].

Information required to manage the organization resources is derived from data. Thus, data is a resource that must be understood, conserved, exploited, employed and integrated. IRM is a new concept, which shifts the design of traditional processing systems around the data to be processed, as the central core. The recognition of the IRM concept, and the awareness of managers that data is an organizational resource, has led them to establish a set of management procedures and technical functions, which called "database administration".

2. Database Administration

Database Administration (DBA) has been staffed from the most of organizations to protect the database, while at the same time to maximize benefits for the users. It is the agency for centralization and exercise of control over the database. Database administrator may be a person, or a group of persons for large, complex, and widely shared databases. It involves several specialties. It includes information system analysts, data structure and data organization designers, security officers, recovery specialists, auditors, and accountants. DBA encompasses all the technical and management activities required for organizing, maintaining and directing the database environment, or otherwise, all the data of the organization which is organized and controlled using a database technology [Ref. 9]. The DBA negotiates with the users; the application analysts, for the storage of data required by the applications; for the permissible use of that stored data; and for the cost/benefit economics and priorities for that stored data. The DBA, the security officers, and the database system maintainers, use the various data descriptive and control languages, and utility programs, to enter policies, procedures, and controls into the database system. The DBA should be the only individual in the organization concerned with the computer's model of the data.

The database environment is a combination of hardware and software that supports the user's interface and the database administrator's interface. It consists of the database, the database administrator, the software tools used in data administration and data processing, and the users. Into this environment the DBA functions as a leader in planning, design, development, implementation, testing, documentation, operation, and maintenance of the entire database. The functions performed by the DBA and the corresponding responsibilities can be categorized into three general areas: (1) management of the data activity, (2) management of the database structure, and (3) the management of the DBMS itself. These functions and responsibilities include: database definition/redefinition; selection and procurement of hardware/ software services; database design/redesign; database creation activities; database security and integrity functions; database maintenance and management; performance monitoring and evaluation; database enforcement activities; liaison activities with users/systems and application analysts; and training of users. A DBA's staff configuration and its its responsibilities is given in Fig. 10. The members have various duties in different stages of the database development. The size of this configuration depends on the size of the enterprise, the complexity of the applications to be run with the database, the complexity of the chosen, the level of the user's sophistication and the phase of the project.

DBA

Operations Expert	2 persons: Expert in communicating with the operations staff, possibly previous operations staff members.
Systems Expert	4 persons: 1 person software expertise major, and hardware minor. 1 person hardware expertise major and software expertise minor. 2 persons experts in DBMS.
Applications Expert	4 persons: 2 persons expert in applications, (possibly previous systems analysts from the application group). 2 persons experts in programming applications (possibly previous application programmers).
Librarian	2 persons: experts in maintaining data dictionary, security definitions, previous experience in user community necessary, articulate in native language.
Query Language Expert	2 persons: experts in formatting screens for on-line transcactions in implementing security, and in authorizations for on-line transcactions, experts in query languages, previous experience in user community.
Auditor	1 person: expert in locating deficiencies in usage of database, in regard to security, privacy, and charging mechanism, on the part of DP operations and application programmers; expertise in accounting, preferably from the user community with some application programming background.

Figure 10.--DBA's Staff at Full Capacity and its Responsibilities [Ref. 10]

3. Database Administration Tools

a. Database Management Systems

A DBMS is a software tool that provides an integrated source of data for multiple users, while presenting different views of the data to different users. It can be considered as generalized software which provides a single flexible facility for accomodating different data files and operations, while demanding less programming effort than conventional programming languages. It features easy access to the data, facilities for storage and maintenance of large volumes of data, and most importantly, the capability for sharing the data resources among different types of users.

DBMS products vary in the degree to which they provide their functions (Fig. 11). Currently, no commercial DBMS provides all these functions entirely satisfactory. These functions are necessary and important however, and this situation should change as DBMS products evolve and as new products are developed.

Although most database experts agree that these nine functions are required, they do not agree how some of these functions should be performed. Some people believe that these functions should be provided by the DBMS automatically. Other believe that some of them should be performed by application programs or by users. In the entire database development cycle, probably nothing is more important than the DBMS evaluation and selection. It is true that DBMS

. Store, retrieve, and update data

. Provide integrity services to enforce database constraints

. Provide a user-accessible catalog of data descriptions

. Control concurrent processing

. Support logical transcactions

. Recover from failure

. Provide security facilities

. Interface with communications control programs

. Provide utility services

Figure 11.--Functions of a DBMS

evaluation is meaningful only after DP managers or DBA have obtained a clear and concise picture about what kind of DBMS will be most beneficial to their particular organization. An intelligence selection of a DBMS can almost guarantee a successful implementation. As many other software evaluations, the selection of a DBMS should include the following items in a checklist with proper weighting factors: Vendor service capability; technical report; personnel training facility; software interface compatibility; hardware requirements; documentation; and security/recovery facility.

The outputs of the logical database design, the system requirements, and the preliminary design of programs, are the inputs to the physical database design. However, since the specific outputs vary from one DBMS to another, we must have an insight of the different database models evolved since their earlier stages.

A database model is a vocabulary for describing the structure and processing of a database. Database models can be used for both logical and physical database design, and used to categorize DBMS products [Ref. 11]. The database models have two major components. The data definition language (DDL), which is a vocabulary for defining the structure of the database, and the data manipulation language (DML) which is a vocabulary for describing the processing of the database. DML are distinguished into procedural DML and non-procedural. Procedural DML is a language for describing actions to be performed on the database. Nonprocedural DML is a language for describing the data that is wanted without describing how to obtain it.

The earliest DBMS were developed in the 1960s, based on hierarchical, network, and inverted-tree data models. Additional improvements led to the currently existing models portrayed in Fig. 12.

Six common and useful database models are portrayed. The models are arranged having an orientation for humans and human meaning, to machines and machine specifications. Their characteristics and application is also shown in Fig. 13. As we can see only three of these models are actually DBMS products.

Human (logical) ←→ Machine (physical)

Semantic Data Model	Entity Relational Model (E-R)	Relational Data Model	CODASYL DBTG Model	DBMS-Specific Model

ANSI / X3 / SPARC

Figure 12--Current DBMS Models

DBMS has significant features and offers advantages in reducing the redundancy of stored data, avoiding incosistencies in stored data, allowing for the sharing of stored data, maintaining data integrity, enforcing standards and providing security over data. Nevertheless, DBMS is one thing, and adopting a database approach is another. DBMS does not necessarily resolve all the problems related to the DBA within an enterprise. In general the hazards of using the DBMS are political and economical [Ref. 12] and can be grouped as:

- Political: Will top management support the creation of database covering the entire organization structure and insist on organization-wide cooperation ?
- Legal: Legal complications can sometimes arise when two or more separate files are joined together into an integrated database.
- Personnel: How many people of what skills must be hired for effective use of the DBMS.

Type	Characteristics	Application
Semantic Data Model	DDL language for storing meaning. High level DML No DBMS based on this model.	Logical database dasign
Entity-Relatio-ship Model	Entities and relationships modeled as data. E-R diagrams graphically show relationships. No DML	Logical and physical database design
Relational Model	Data represented as tables. Relationships implied by data. Dynamic data relationships. Procedural and nonprocedural DML. A few DBMS based on this model.	Logical and physical database design
CODASYL DBTG Model	Oldest data model. Relationships must be predefined. Procedural DML. Extensive application in industry. Many DBMS based on this model.	Physical database design
DBMS-Specific Model	Models vary widely. DDL and DML closely conform to features of the DBMS.	Physical database design
ANSI/X3/SPARC Data Model	DBMS model instead of database model. Three schema model. Can support a variety of different data models.	Design model of DBMS

Figure 13.--Characteristics and Application of Data Models

- Training : What training courses must be provided to prepare personnel for database development and use ? Training of user personnel must be given as much consideration as training of the developers.

- System software : What other software packages must be purchased for effective use of the DBMS.

- Application software : All database implementations require the development of application software. Software aids provided by the DBMS to reduce the skill level and time required to develop application software is important to the overall cost of DBMS use.

- Hardware : What additional hardware must be added to make effective use of the DBMS.

b. Data Dictionary / Directory Systems

A successful enterprise is the mirror of an effective management. And a vigorous management is cognizant of the value of corporate resources, and optimizes and safeguards them accordingly. Until quite resently only a small percentage of enterprises exercised as tight a control over their data resources as over their other resources. And as data resources continue to increase, it is becoming apparent that the DBMS is not the panacea to all the problems about the management of data, particularly since not all the data is automated. DBMSs encouraged centralization of perspective on how data should be organized and used. A recent trend is to use a new automated tool which encourages centralized perspective on how information about these data should be organized and used. This tool is called Data Dictionary/ Directory System (DD/DS). The DD/DS performs some of the same functions as the DBMS, and provides benefits parallel to those attributed to the use of a DBMS. However, it has begun to be recognized as the more general of the two, since the benefits derived from its use are related to the total data resources of an enterprise. It provides the central control of data resources in a uniform manner across organization lines. It pertains not only to database systems but increasingly to non database systems as well, and it is further broadened by its support for job streams, structured systems, on-line environments and system design. It may be considered as the primary software tool for a DA, providing the possibility in an enterprise with no need for a database administrator (DBA).

c. Data management Software Tools

There are a great number of commercial software data management products available today, which can assist the DA, or the DBA in an enterprise. These products can be classified into four general categories. Those based on physical linkage between files, those based on inversion of data values, the decision support systems (DSSs), and the file-pass data manipulation systems (DMSs). The characteristics of these management software tools, and the current offerings by vendors are shown in Fig. 14-15.

Ross [Ref. 13], categorizes these four types of management systems, into three conceptual classes : Designer packages, self-contained packages, and implementation and

	Designer Packages	Self - Contained Packages		Implementation and Access Tools
System Characteristics	Physically Linked DBMS	Inverted DBMS	DSS	File-Pass DMS
Solely responsible for data management	Yes	Yes	Yes, but data often extracted other sources	No, data management support usually provided by another method such as ISAM, or a DBMS
Structural flexibility and capacity for evolution	Low	Moderate to high	Moderate to very high	Does not apply, since does not provide its own structural support
Degree of system support for data integration	High	Moderate to high	Integration is often not necessary or not considered a primary object	Little or no assistance in this area
Sharability of data between users, including those performing updates	High, including updating	High, including updating	Variable	Minimal or none, especially for updating
On - line orientation, especially for multiple users	High	High	Variable	Minimal or none
Recoverability, including selective program back-out	High	High	Variable	Does not apply

Figure 14.--Classification of Commercially Available Data Management Software [Ref. 13]

Capacity for performance tuning	High	Low to moderate	Low	Minimal
Package orientation	Systems staff	Mixed; end users and systems	End users, especially managers	Usually considered a programmer's tool
Ad hoc data access capacity	Minimal	Moderate to high, via query language	High, between either query language or reporting facility	High, via reporting facility
Performance on repetitive production tasks (assuming optimum database design)	Moderate to high	Moderate	Low to moderate	Low to moderate
Characteristic data volatility	Moderate to high	Moderate	Moderate to minimum volatility (upon reaching decision-support system)	Low (when file-pass system in use)
Data usage orientation	Production type data	Production oriented with ad hoc query potential	Oriented to analytical, statistical, and/or searching users of data	Variable (batch reporting of variably oriented files
Typical structure or organization inherent in the data managed	Data highly organized and structured	Data moderate to highly organized and structured	Capacity for handling low organized data (e.g., textual or numerical data	Does not apply (data structuring is usually file-oriented, but may be more under a DBMS)
Overall complexity of system	High	Moderate	Low to moderate	Low

Figure 14--Classification of Commercially Available Data Management Software (continued)

Designer Packages	Self-Contained Packages		Implementation and Access Tools
Physically Linked DBMS	Inverted DBMS	DSS	File-Pass DMS
IMS/VS (DL/1) IBM	ADABAS Software AG	SYSTEM 2000 Intel	MARK IV Informatics
TOTAL Cincomm Systems, Inc.	DATACOM/DB Applied Data Research Corporation	INQUIRE Infodata Systems, Inc.	
IDMS Cullinane Corporation		RAMIS II Mathematica, Inc.	ASI-ST Applications Software
DMS-II Burroughs		DMS-170 CDC	
I-D-S/II Honeywell	Model 204 Computer Corporation of America	SYSTEM 1022 Software House, Inc.	Data Analyzer Program Products
EDMS Honeywell		MANAGE Computer Sciences Corporation	CULPRIT Cullinane Corporation
DMS 1100 UNIVAC		MAGNUM Tymshare	CULPRIT Cullinane
DMS/90 UNIVAC		NOMAD National CSS, Inc.	
DBMS-10 DEC		DPL National Information Systems Inc.	
		FOCUS Information Builders, Inc.	
		ORACLE Relational Software, Inc.	

Figure 15--Classification of Commercially Available Mainframe Data Management Software [Ref. 13]

access tools. The examination of the various system types reveals that the physically linked DBMSs are highly complex, but the benefit of their use is a high degree of performance tunability. These packages are commonly oriented toward DP professionals, like the application designers. The inverted DBMSs and the DSSs, are much less complex, more flexible in meeting application requirements and can be characterized as self-contained systems. Finally, the file-pass DMSs, rely on some other system component (an access method, such as ISAM) or DMSs, facilitate the task of data delivery by circumventing the need for application programming, and labaled as tools for program implementation and data access.

G. SUMMARY

The introduction of the computer technology, and the proliferation of its usage into organizations, led them at ealier stages to believe that all their problems could be solved by machines. Later, they realized that most of the problems were not technical, but rather administrative and procedural. Thus, many concepts such as, database, IRM, were developed and implemented for more efficient operation and control.

Data has been an important element in the operation of an organization throughout history. Nevertheless, it is not until recent years, with the explosive proliferation of computers, that the value of data as a resource has been fully recognized.

The human function responsible for the proper administration, control, and coordination of all data-related activities into an organization is the DA. A primary software tool, an automated facility, that supports the data administration function in managing data as a resource is the Data Dictionary/Directory System.

III. DATA DICTIONARY/DIRECTORY SYSTEMS

A. DEFINITIONS AND BASIC PRINCIPLES

There is a wide range of definitions for Data Dictionary /Directory Systems (DD/DS). Due to the increasing interest and the rapidly evolving nature of this field in recent years, terminology is somewhat confusing. One author speaks of a Data Dictionary System (DDS) or System Resources Dictionary, while another refers to DD/DS or Data Element Dictionary/Directory System, (DED/DS). Characteristic definitions include those of:

Leong-Hong and Marron (1977):

> the DED/DS is a software tool that provides the means for defining and describing the characteristics of a database, as opposed to the contents of a database [Ref. 14];

National Bureau of Standards (NBS) (1978):

> the DED/DS is considered as a resource manager. It is an integrated repository that provides data necessary for managing data. Data management includes the planning, control, direction, and organization of data [Ref. 15];

Cardenas (1979):

> the DDS is a centralized repository of data about data [Ref. 16];

Leong-Hong and Plagman (1982):

> the DD/DS is a system that is designed to support comprehensively the logical centralization of data, about data (metadata), [Ref. 17];

Further, they elucidate the difference between the two types of metadata, "dictionary" and "directory". Dictionary metadata provides information which describes what the data is what it means, and what exists. Directory metadata describes where the data is located, and how it can be accessed.

Van Duyn (1982):

> the DD/DS is a collection of data elements, structure informational entities, characteristics, and locations of data, [Ref. 18].

He also, proposes that the current concept of a data dictionary system is composed of the following:

> Data Catalog: a structured listing of data elements, with or without a description of the listed elements.
>
> Data Dictionary: an organized compilation of data elements, data attributes, structure, and characteristics.
>
> Data Directory: an orderly listing of data elements names, identifiers, locations and physical characteristics of these data.
>
> Data Dictionary/Directory: which combines the features of a data dictionary and data directory.

Allen, Loomis and Mannino (1982):

> a DD/DS is an automated information system. It helps to achieve control of the data resource, by providing an inventory of that resource. It helps to control the cost of developing and maintaining applications. Finally it can provide for independence of metadata across computing environments, improving resiliency to the effects of hardware and software changes, [Ref. 19].

The variety of definitions provides some idea of the evolving scope and increasing complexity of DD/DS. The terms "data dictionary", "system resources dictionary", "directory of data definitions", appear in common usage. The DD/DS does not contain only definitions for data, nor it is merely a dictionary. It is not only a "system resources dictionary", because many DD/DSs contain corporate information about users and procedures, that are not really system resources at all.

Unfortunately, in spite of these shortcomings, no better designator has yet been suggested, and the terminology remains confusing. Part of this problem, is due to the fact, that DD/DS technology has evolved very rapidly in recent years. Also, since DBMSs were developed before DD/DSs, there is a natural tendency to view DD/DS as subordinate to DBMS. Especially, in cases where a DD/DS-like function is part of the DBMS, or where it cooperates with a DBMS. Nevertheless, the increasing interest and improvements in DD/DS, and the development of DD/DS independent from DBMS, has caused it to evolve into a complex software tool which should be considered more efficient than the DBMS.

In this thesis it is assumed better to describe the features and functions of a DD/DS, rather than to precisely define what the DD/DS really is. References to DD/DS will imply that dictionary and directory functions are available, unless specifically stated otherwise.

The basic principles that will serve as a foundation in further discussion include the following:

- Data has positive or negative value. The value of the data derives from the fact that the entire enterprise depends on its availability for the proper management of all other resources. Thus, it must be treated as a resource. The management and control of data resources begins with a proper definition and description of data. A DD/ds is a tool for the control and management of data as a resource.

- To manage data as a resource it is basic that data about data must be clearly specified, easily accessible and well controlled. These data are called metadata. These are data objects, that in a data processing environment are represented in the form of elements, records, files or databases. Metadata is not user data, but identifies, defines and describes the characteristics of the latter. It describes the data resources of an organization. A DD/DS contains two types of metadata: Dictionary and Directory metadata. Dictionary metadata describes the data, and defines their meaning and structure. Directory metadata describes where the the data is stored, and how internally represented and accessed.

- A collection of related metadata comprises the metadata database [Ref. 20]. It consists of a database that contains descriptive and definitional information about the user database. It has basically the same characteristics of a user database. To achieve the goals of managing data as a resource it requires proper management. That is, planning for the design, implementation, maintenance, utilization and control. This implies that established lines of responsibility and authority; formal rules and detailed procedures to guide metadata-related activities; common procedures for collection, update, and maintenance; and common procedures for access control to the metadata must be developed. The DD/DS is the basic tool for managing the metadata database.

- The DD/DS is divided into three categories, based on the scope of control exercised through metadata management: Active, Potentially Active, and Passive, [Ref. 21]. A DD/DS is said to be active with respect to a program or process if that program or process is fully dependent on the DD/DS for its metadata. A DD/DS is said to be passive if it does not generate metadata and does not have control over where and how a user or processing component obtains the required metadata. A potentially active DD/DS provides the capability of producing the metadata for a given program or process. A potentially active DD/DS can be extented to active through supportive administrative procedures. Many of the currently commercially available DD/DSs are of this type. In practice, the concept of active/passive DD/DS refers to interfaces that it provides to other software packages. A DD/DS with active interfaces can better serve the goals of managing data as a resource.

B. FEATURES OF DD/DS

The underlying design philosophy of a DD/DS is that it will serve the needs of a wide variety of users in an organization. Six groups are identified that can/should share the DD/DS:

- Data Administrators/Database Administrators (DA/DBA), who use the system to control the data resource, iplementing standards, designing, monitoring, and restructuring.

- System Analysts and Programmers, who share the DD/DS to facilitate system design and system implementation activities.

- Operations Staff who retrieve information about jobs.

- Data-Processing Management, who receive the high-level impact and summary reports about data usage from the DD/DS.

- End-Users, who access the DD/DS to obtain information about existing data and system resources.

- Auditors, who examine system documentation provided through the DD/DS.

Each of the users will contribute metadata to the DD/DS, having different needs, and different logical views of the metadata. The necessity to support such diverse logical views of metadata among users requires the database approach in the design of a DD/DS. This, includes three steps: (1) identify the entities (meta-entities), (2) define them, and (3) identify and describe their relationships.

Allen, et. al. [Ref. 22], delineates the logical structure of a typical DD/DS's database. The DD/DS is represented by a network data model of entities, relationships, and attributes. An entity is an object of interest about which information is collected. A relationship is an association between one or more entities. An attribute is a characteristic relating to an entity or relationship. Attributes which can be used for an entity or a relationship in a DD/DS are: type, range, length, unit of measure, usage, language names, repetitions, keys, defaults, and display formats. It is essential that the structural characteristics of the DD/DS be described in logical terms. This will provide a clearer insight into what kind of metadata are supported by it. The relationships between entities of a typical DD/DS are illustrated in Fig. 16. The user entity although not shown may be related to nearly all the other entities. Thereby, representing the user-subschema, user-process, user-terminal, user-transaction, and user-report relationships.

A recommended classification of entities (meta-entities) and attributes, [Ref. 23], is also listed in Fig. 17,18. The entities are classified into three groups: (1) data entities (2) system or processing entities, and (3) environmental entities. The attributes into seven groups: (1) identification, (2) representation, (3) statistical, (4) relationship, (5) control, (6) physical storage media, and (7) user-defined attributes.

The last category of user-defined entities, relationships and attributes is one of the most important features of the DD/DS. It is the expansibility or extensibility

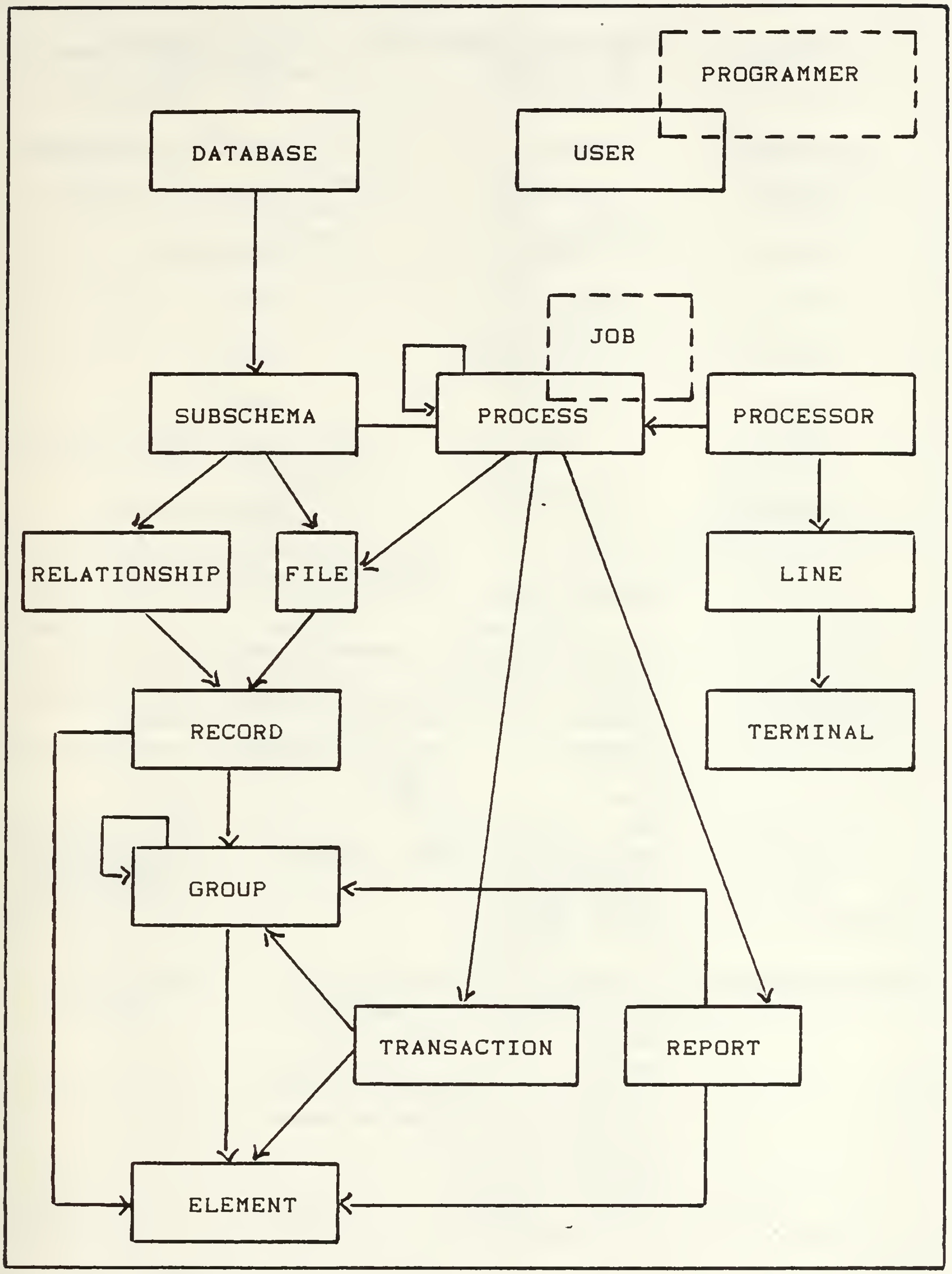

Figure 16--Logical Structure of a Typical DD/DS [Ref. 22] with Expansibility Feature.

Data Entities	System/Processing Entities	Environment Entities
Element	System	Physical devices
Record	Subsystem	Computer system
Group	Program	Terminal
File	Module	Line
Database	Transaction	Users
Screen		Node
Report		Function
Subschema		Organization

Figure 17--Classification of Meta-Entities [Ref. 23]

Category of Attributes	(Data) Element	(Process/System) Program	(Environment) User
Identification	Name ID Description Synonym COBOL-name Alias	Name ID Description	Name ID Description Description of position
Representation	Character type Source type Length Picture clause Format Justification Content type Initial type	Programming language Number of parameters Segmentation Type of code	Function code

Figure 18--Entities/Attributes Matrix [Ref. 23]

Category of Attributes	(Data) Element	(Process/System) Program	(Environment) User
Statistical	Frequency of access Access type Volatility	Performance status Log information Access type Processing type Overlay indicator	Usage code read only, update, etc
Relationship	Keywords Occurs in Indexed Used in Appears in Relationship	Keywords Calls subroutine	Creator User Modifier Maintainer Related person
Control	Password Encryption Version Status Origin Destination Value range	Password Encryption Version Status	Responsibili-ty code Level of security Authority
Physical Storage Media	Storage media	Storage location Storage media Storage size Compiler Expected CPU time CPU Procedure Required peripherals	Authority Contact-phone Mail address
User-defined	User-defined	User-defined	User-defined

Figure 18--Entities/Attributes Matrix (continued)

feature. It is the capability, in a more dynamic approach, that allows a user to expand the vendor standard meta-entity structure, to suit specific corporate needs. Extensibility also, can be provided by vendors by optional sets of data objects. This feature becomes increasingly important as it evolves into a means for coordinating software development at each stage of the application life cycle. User-defined entities are sometimes subject to constraints, such as, a new entity must assure the relationships of a predefined entity. To use the extensibility feature of a DD/DS effectively three common-sense rules are applied:

- Maintain consistency in the design of the metadatabase.
- Maintain a proper balance in the definition of entities and attributes.
- Use clearly defined, unambiguous attributes.

Another important feature of a DD/DS is the ability for entity occurences to exist in different states such as test, production, and historic [Ref. 24]. Test status information is used to document evolving systems and uncopleted changes to production systems. Production status occurences in the DD/DS are reflections of operational systems. Finally, historic status reflects superseded metadata, thus, providing an audit trail of changes to the DD/DS.

A DD/DS includes one or more interfaces that allows the user to interact with the DD/DS, as well as interface facilities to other systems. Such interfaces include:

- Command languages (maintenance, report and query, data structure interface commands, extensibility and status-related commands, security commands, processing control commands, and administrator commands).

- Screen oriented interface.

- A fixed format batch data entry facility.

- Interface that allows user written application programs to access the DD/DS.

- Support interface to Report Writer and Query Systems.

- Edit and validation criteria.

- Test data generation.

- Application development aids.

These interfaces depend on the nature of the DD/DS. The contents of a DD/DS can be extremely broad, as its basic structure lends itself to the control of a wide variety of processes.

A final feature is the DD/DS control and security. Control in a database environment can be defined as:

> the plan of organization and all the coordinate methods and measures adopted by an enterprise to safeguard its assets [Ref. 25].

Check the accuracy and reliability of the data contained in the database, promote operational efficiency and encourage adherence to prescribed managerial policies. Control is a positive force designed to direct an operation to its successful completion. The DD/DS function is essential in controlling the consistency and use of data resource.

The primary control issues that a DD/DS addresses are the following:

- Data definition control through proper documentation.
- Data usage control through use of a proper accounting subschema.
- Enforcement of standards control through facilities enforcing standards.

When an organization acquires its own DD/DS system and depends on it, the continued proper operation of that system is important. The losses which follow from failure can be considerable. Security and privacy in the operational level requires that a robust and correct system is in use and understood by all involved, and that adequate enforcement of the system exists. The security and privacy issues in a DD/DS concern:

- Access to the system (prevent unauthorized access, and permits authorized access).
- Damage to data, software and equipment.

Most manufacturers do supply some security and privacy controls as part of the operating system or the DBMS. Security mechanisms can be provided by the DD/DS, but remain relatively unexplored today.

C. FUNCTIONS OF DD/DS

The typical functions performed by a commercially available DD/DS are listed in Fig. 19. These functions are:

- DD/DS Maintenance:
  interprets and processes requests to add, change or delete contents of the DD/DS
- Extensibility
  enables the DD/DS structure to be extended by the definition of additional entities, relationships, and attributes.

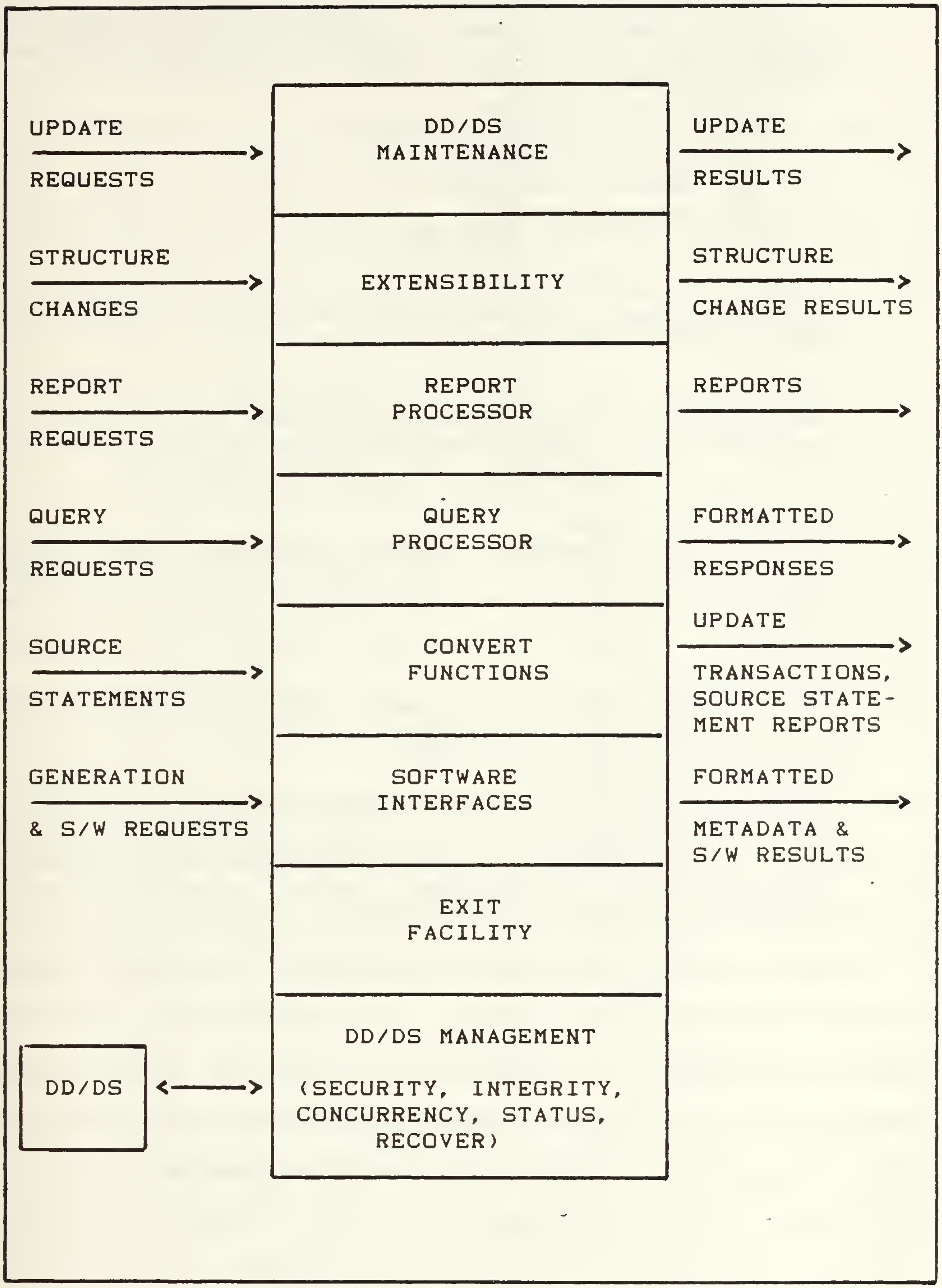

Figure 19--DD/DS Functions [Ref. 26]

- Report Processor:
provides predefined reports, the ability to customize reports and user-defined reporting capability. Common predefined report types include:
    (1) name listings
    (2) relationship reports
    (3) detail reports
    (4) summary reports
    (5) matrix reports
    (6) graphical reports

- Query Processor:
allows English-like queries of the DD/DS (used for low volume retrievals).

- Convert Functions:
reads application programs, libraries, and schemata and generates input transactions for the DD/DS Maintenance Function which describes the detected metadata.

- Software Interface:
provides a formatted pathway enabling the DD/DS to provide metadata to other software systems and enables these systems to retrieve and update information in the DD/DS.

- Exit Facility:
enables the vendor-supplied routines to be extended (not available in all DD/DS).

- Database Management:
performs database management tasks. Security, integrity concurrency control, and internal access for the database. This function is often performed by utilizing an existing DBMS, nevertheless, DBMSs do not provide all necessary subfunctions of this function.

These software interfaces, and convert function capabilities, and permit the DD/DS to be intergrated with other software packages. The facilities used for that purpose allow direct and indirect access to the DD/DS; and automatically capture the metadata used by other systems.

D. DD/DS A TOOL IN SDLC

The DD/DS is a software tool originally intended as a documentation aid for data management. The software documentation feature became the primary goal of developing DD/DS in the mid 1970s. Since then the spectrum of applications for which DD/DSs have been used has been wider. DD/DSs have been found useful in many areas of computer processing and data resource management.

The British Computer Society (BCS) established a study group, the Data Dictionary System Working Party (DDWP) in January 1975. Over a period of time the BCSDDSWP worked to produce a report on the need for and the facilities which should be provided by a DD/DS; and related database design and operational units. They studied the currently available DD/DSs and the related design aids. They identified data recording and analysis needs for the design of information systems. They specified requirements for aids to database design, maintenance and operation, and considered which of these requirements were automatable. The report of this study group was published in late 1977. This study emphasizes the shift to expanding the functions of a DD/DS from a software tool with cataloging the data in an existing database, into an adjunct to design the database itself. This study also emphasizes the use of a DD/DS throughout the complete specification, design and implementation stages of the SDLC. Particular functions which could be performed would be:

- data analysis, to determine the data structure;
- functional analysis, to determine the way in which events and functions use data;
- database or conventional file design;
- storage structure design;
- operational running of the application systems;
- collection and evaluation of performance statistics;
- database tuning, to improve performance;
- application maintenance, and database restructuring.

The BCSDDSWP Report further recommended that the DD/DS should provide two sets of facilities: one set "the conceptual data model", would record and analyze requirements independently of how they were to be met; the second set the "implementation data structure", would record design decisions in terms of the database or file structures implemented, and the programs that would access them [Ref. 27]. For both facilities it is necessary to record data usage, as well as, data structure, giving rise to four areas of information which can be identified. Fig. 20 depicts these four areas.

The DD/DS should relate definitions of the implementation data structure to the parts of the conceptual data model they describe. Recording the mapping documents, design decisions, and clarifies the decisions which have been made.

The first stage of any Information System Development Life Cycle is the Planning Phase or the Perception of Need Phase. The purpose of this activity is to determine the feasibility, technical, and economic trade-offs for a planned

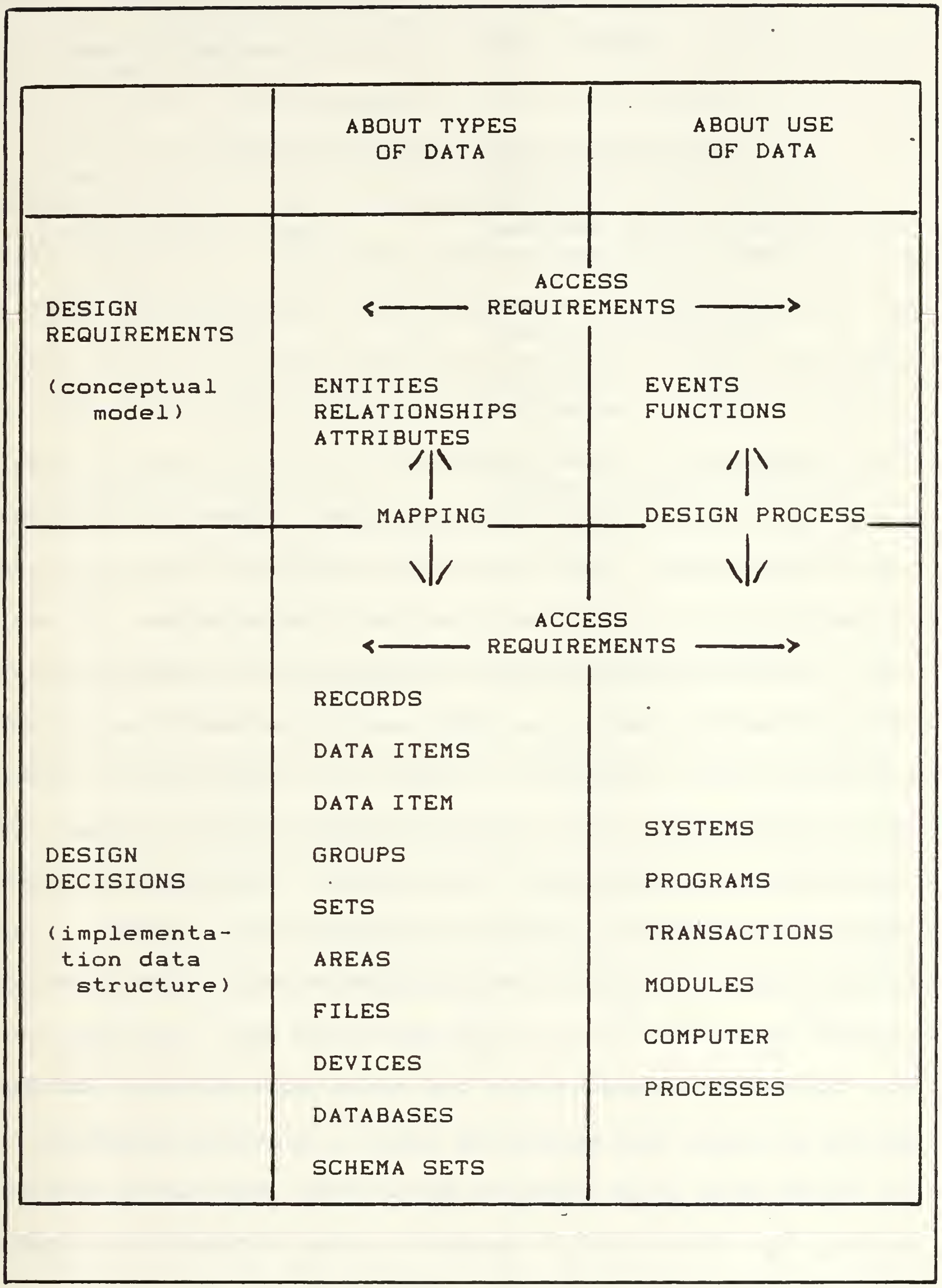

Figure 20--The Information in a DD/DS [Ref. 27]

system. It is strategic in nature, a continuous process throughout the active life of an Information System. This system may already exist in the organization, requiring some new kind of functionality, or may be a new one.

The planning process begins with an initial assessment of the current environment, and continues with an analysis of current usage, and determination of future requirements. These activities determine the data needed; the data that is already available; the potential conflicts and redundancies; and the impact on existing systems, and the potential users. By their nature they require a high degree of consistency and coordination. The DD/DS can be used to support these activities , as a documentation tool, as well as, a design aid. Information about the data usage, their relationships, and attributes, can assist the analyst in recording the flow of data across functions. Conflicting usages can be identified and resolved, and redundant data can be removed from the database. Analysts can also use the DD/DS to predict the impact of a proposed change and define what actions should be taken to prevent unwanted side-effects. Analysts must consider the DD/DS as a comprehensive tool for collecting all the facts necessary for the clear and complete statement of the problem, and providing data to test the solution.

Leong Hong and Plagman [Ref. 28], recommend the following use of a DD/DS in support of the information systems planning and modeling process (Fig. 21).

Steps in Information Systems Planning	Use of DD/DS
Definition of business function	Document business functions and input/output
Definition of data clusters	Document data clusters and identify redundancy
Definition of data cluster usage	Document usage and perform analysis
Definition and analysis of transactions	Document transactions and perform analysis
Development of a conceptual data model	Document data structure and perform impact analysis

Figure 21--DD/DS Support for Systems Planning [Ref. 28]

The second stage is the Requirement Definition and Analysis Phase. This phase describes how the "real world " operates. What needs to be done to achieve the predefined goals. What kinds of reports are needed; what activities will produce these reports; and where are the sources of the needed information? Requirement Definition and Analysis can be a difficult process that requires manipulation of the business functions and of detailed conceptual model. The use of a DD/DS in that phase facilitates the documentation of the requirement definition and supports the analysis.

Lefkovits, et. al. [Ref. 29] identifies four very important results, that derived from the use of a DD/DS in this phase:

(1) The requirements that are specified can be analyzed to assure that the system being defined meets the

objectives associated with the strategic, tactical, and operational management of the enterprise. The importance of a DD/DS is that it can be used to store the different views of data which are associated with these three levels of management. Additionally the data that is specified as part of the requirement definition can be validated against the DD/DS. Thus, the entire enterprise and its functions can be modeleded in the DD/DS.

(2) The requirements in the DD/DS can be analyzed for internal consistency. This solves the problem of conflicting requirement specifications.

(3) The DD/DS containing the description of the system to be changed, facilitates the decision on the cost effectiveness of the modification.

(4) Finally, the availability of the requirements in a well-organized and processible manner will tend to improve the quality and efficiency of the tasks that must be performed in the succeeding stages.

The design phase, the third stage, is a critical phase. This is due to the fact that management-after another review of justification for the undertaking- will postpone the project. Normally, because only a general or overview design of the project is prepared. The objectives of the design phase are intended to find a "how-to" solution. This phase is possibly divided into logical and physical parts, to allow logical aspects to be defined before any implementation considerations might affect the more abstract considerations. During this phase an implementation data structure is also constructed. Use of DD/DS during this phase facilitates modeling of data structures and the process of database design

The DD/DS should be used in the logical and physical design. It stores the descriptions of the system components, such as, subsystems, program modules, data structures, data access techniques, and data flow. It documents the logical and physical design, by describing the data required by the programs. It provides flexibility by allowing individual findings or decisions to be recorded in the appropriate places in the DD/DS. It provides reference to any item of data of interest to the individual. Multiple user views can be recorded in the DD/DS. Multiple designs can be generated efficiently for performance testing and simulation. Conversion of data can be verified. Further, the DD/DS should be used actively to assist, by generating the DBMS control blocks from the metadata.

The Implementation and Test phases deal with the collection of data, coding of programs, testing of both data and processes, and overall system debugging. The DD/DS is one of the few tools that provides any degree of coordination and control over the tasks that are performed in this phase. During input to the implementation model, the systems analyst can call on the DD/DS to perform consistency checks to ensure that the new data is complete and correctly formatted. Validation by the DD/DS can confirm proper mapping between the data model and the implementation model. Since the DD/DS -monitored by the DA- contains information as to who may have access to specific data sets, records, segments etc., access control also is gained. Finally, the DD/DS in this phase is very useful when it can generate metadata; including a data division for a COBOL program, or a schema for a DBMS, or the required metadata for any other software component.

During the Operation phase, next stage, there may be a need to refine the system to improve its performance. The DD/DS can be used as a tool to support a smooth transition between the new and the old system; or to support a smooth start up of a new system. It can store instructions for the operational staff in many areas: descriptions of various activities; instructions for actions to be taken in case of abnormal termination of a process; statistics about the operation of program in the system, etc.

The Maintenance phase is the last phase in the SDLC. It is the phase where reorganizations or modifications are taking place. Some authors argue that this stage is perhaps a total or partial iteration of the entire SDLC; and this is true. Using the DD/DS appropriately in the SDLC we can maintain any system under development or operation.

Fig. 22 summarizes the SDLC functions which can be supported using the DD/DS features.

E. BENEFITS OF USING DATA DICTIONARY/DIRECTORY SYSTEMS

A DD/DS impacts the management of an organization by improving management's control and knowledge of the data resource. This control and knowledge is achieved using a software tool--the DD/DS. The benefits that are derived vary

SDLC Phase	SDLC Function	DD/DS Feature
System Planning	Collect information about the organization's data needs	Definition service Cross-referencing Relationships and dependencies Keyword analysis Impact analysis
Requirements definition and analysis	Analyze feasibility, cost/benefit, impact analysis	Impact analysis Cross-referencing Usage Statistical summaries
Design	Specify functions to be accomplished Modeling of structures Design schema	Define detailed characteristics of entities Version control facility Impact analysis report Change analysis
Implementation	Generating schema DDL, COBOL data division Documentation of the database	Metadata generation for DBMS, or COBOL
Test	Produce test data	Support generation of test data
Operation/ Maintenance	Changes to entities Reorganization Restructuring Conversion	Change analysis and control Test and production facilities

Figure 22--The Use of a DD/DS in the SDLC

from one organization to another. In a database environment with a larger number of data elements, relationships, and relatively high number of changes, the benefits with a DD/DS will be higher.

These benefits also, seem to increase as the sharing of data elements between programs increases.

Some of the benefits using a DD/DS in an enterprise involve the following aspects:

- Enables management to enforce the data definition standards.
- Minimizes unwanted data redundancy.
- Assists in securing sensitive data.
- Assists management in quickly determining impacts of proposed changes to a system.
- Provides better data integrity than a file management system or any variety of DBMS.
- Assists management in ensuring complete and accurate changeovers in the implementation of new systems.
- Provides information about the creation, usage, and relationships of data.
- Reduces the clerical load of a DA/DBA.
- Gives a DBA control over design and use of a database by:
  - (1) controlling and documenting formulation, meaning and usage of data structures;
  - (2) evaluating and controlling data redundancy; and
  - (3) providing accurate data definitions for programs
- Supports in the analysis of an organization's data flow by providing a method to track documents which flow through an organization.

- Provides a central source of information for the designers and analysts to prevent data redundancy and incosistency.

- Generates test data for designers.

- Provides documentation on systems design.

- Enforces data definition standards during program coding.

- Automatically generates code.

- Improves the accuracy of finished programs by generation of test data and checking results automatically.

- Provides cross-referencing to assist in implementing approved changes to a system.

- Implements automatically amendment to operational systems.

- Provides documentation on changes to a system.

- Aids operational personnel in the creation of JCL parameters.

- Determines the source of data (including invalid data).

- Aids the auditing function.

- Interfaces to application program development tools.

The above mentioned benefits may be considered as tangible benefits. One last benefit that DD/DS provides does not have a tangible value. It is the benefit that is derived from a properly implemented and well managed DD/DS. The trustworthy information for all users in an enterprise.

IV. COMMERCIALLY AVAILABLE DD/DS PACKAGES

A. OVERVIEW

In this chapter most of the currently, commercially available DD/DSs are discussed. The intention is to reveal the differences between theory, promise, and practice. The major problem is a lack of a standard methodology which ties a DD/DS into the system life cycle.

The available commercially DD/DS packages can be grouped into four categories:

(1) Independent DD/DS:
a free-standing package which can be used in a non-DBMS environment.

(2) Independent DD/DS with interfaces to other DBMSs:
a free-standing package that optionally provides interfaces to one or more DBMS.

(3) Dependent DD/DS:
the software package is designed to co-exist with a particular DBMS. It is dependent on the DBMS and used as a front end process to the particular database or DBMS.

(4) Embedded DD/DS:
a fully intergrated DD/DS with the DBMS. It can enhance the data control and management capabilities of a particular system. The DD/DS function usually is part of the data definition function, and the meta-data is stored as part of the database for the DBMS.

The principal advantages of the independent DD/DS are the following:

- They are self-contained, performing their functions independently of any database(s) or DBMS.

- They can store computer data, as well as, nonmechanized data.

- There is less risk involved of commitment to a database management system in implementing an independent DD/DS than a dependent one.

- They do not need all the data descriptions required for a database at one time, as the dependent one.

- They can support one or more DBMSs through their interfacing capability.

The principal advantages of the dependent or embedded DD/DSs are the following:

- They provide information as to the location of data within the physical database.

- They can serve as a much more powerful control tool when integrated with the DBMS, because the database designer and the users will have to enforce the DD/DS as a tool for documentation and control of the data.

- They provide data validation through embedded range and criteria checking.

- Technical software coordination issues between a dependent DD/DS and a DBMS are minimized.

- The selling effort to current DBMS users is easier.

- The development effort for the DD/DS is much easier.

Seventeen systems were selected based on a survey of the literature [Ref. 30, 31, 32]. The criteria that were used primarily concern the following areas:

- Type of DD/DS.

- Language used.

- Entity names used.

- Expansibility features.

- Status capability.

- User-defined reports

- Security levels provided.

. Software interface packages and methods.

Fig. 23 contains the selected areas of interest. These DD/DSs are designed especially for large mainframes.

SYSTEM NAME	VENDOR	FIRST RELEASE/ LAST RELEASE	CATEGORY
DATA CATALOGUE 2	Synergetics Corp.	1974/1977	Independent
DATAMANAGER	Management Systems and Programming (MSP)	1975/-	Independent
PRIDE/LOGIC	M. Bryce and Associates	1974/-	Independent
LEXICON	ANDERSEN A. and Co.	1972/-	Independent
EDICT	Infodata Systems Inc.	1972/ in development	Independent or DBMS Application
TIS DIRECTORY	Cincom Systems Inc.	1979/ in development	Embedded
ADABAS DATA DIRECTORY	Software AG of North America, Inc.	1978/ in development	DBMS-Application
DATA CONTROL SYSTEM	Cincom Systems Inc.	1976/1980	DBMS-Application

Figure 23--Features of Commercial DD/DS Packages

SYSTEM NAME	VENDOR	FIRST RELEASE/ LAST RELEASE	CATEGORY
DATA CONTROL SYSTEM (DCS)	Harvey Systems, Inc.	1976/-	DBMS-Application
DATA DICTIONARY SYSTEM (DDS 1100)	Sperry Univac	1981/-	DBMS-Application
DATA DICTIONARY SYSTEM (DDS)	International Computers Ltd. (ICL)	1977/-	DBMS-Application
DATA DICTIONARY	Applied Data Research Inc. (ADR)	1979/-	DBMS-Application
DB/DC DATA DICTIONARY	IBM	1974/-	DBMS-Application
DICTIONARY/204	Computer Corporation of America	1982/-	DBMS-Application
Integrated Data Dictionary (IDD)	Cullinane Corporation	1977/-	DBMS-Application
Extended Data DICTIONARY (XDD)	Intel Systems Corporation	1970/1980	DBMS-Application
UCC TEN	University Computing Center (UCC)	1970/-	DBMS-Application

Figure 23--Features of Commercial DD/DS Packages (continued)

SYSTEM NAME	SOURCE LANGUAGE	ENTITY NAMES
DATA CATALOGUE 2	COBOL	Element, Group, Record, Resource, Form, Task, Report, File, Module, System, User
DATAMANAGER	Assembler	System, Program, Module, Database, File, Group, Item, Segment, PCB
PRIDE/LOGIC	COBOL	System, Subsystem, Procedures, Programs, Modules, Files, Inputs, Outputs, Call arguments, Records Data elements
LEXICON	Assembler and COBOL	Data elements: Item, Subgroup, Group, Record, Entry, Database, File, Sensitivity Processing entities: Validator, Program, System
EDICT	Primarily PL/I	Element, Database
TIS DIRECTORY	Assembler	System data: Component, Reserved word, Mask, Edit, Translate table Schema data: Environment, File, Environment file, Buffer pool, Internal record, Physical field, Subschema, Access set, External field User data: User, Procedure, Expression, Equation
ADABAS DATA DIRECTORY	Assembler, COBOL, NATURAL	Fields, Relationships, Databases, Files, Field verification, Procedures, Owners/Users, User views, Programs, Modules, System, Reports, Responce codes
DATA CONTROL SYSTEM (Cincom)	COBOL and MANTIS	User, Report, Program, Document, System, Source, Element, Database, File, Transaction, Physical/ Logical Element
DATA CONTROL SYSTEM (DCS)	COBOL	Schema, Set, Area, Subschema, Record, Field, Program

Figure 23--Features of Commercial DD/DS Packages (continued)

SYSTEM NAME	SOURCE LANGUAGE	ENTITY NAMES
DDS 1100	PLUS	Module, Run, Unit, DBA, Analyst, Application, Sshema, Subschema, Area, Record, Group, Data-item, Data-name, Set, Database, Stream, Procedures, Run, File, Record, Relation
DATA DICTIONARY (DDS)	COBOL	Entity, Relationship, Attribute, File, Virtual file, Record, Group, Item, Operation, Event, System, Program, Module, PMAP, DMAP, Area, Schema, Subschema, Set
DATA DICTIONARY (ADR)	Assembler	Database, Area, File, Key, Element, Record, System, Person, Job, Step, Authorization, Module, Program, Report
DB/DC DATA DICTIONARY	Assembler	Database, Segment, Element, PCB, SYSDEF, System, Job, Program, PSB, Module, Transaction, DDUSER
DICTIONARY/ 204	Model 204 user language	File, Group, Record, Field, Procedure
Integrated Data (IDD) Dictionary	Assembler	User, System, Program, Entry, Point Module, Element, Record, File, Task Queue, Map, Panel, Line, Physical terminal, Logical terminal, Destination, Message
XDD DATA DICTIONARY	Assembler	Application, Work unit, Program, File, Work area, Work structure, Database, Schema, Subschema, Item, User
UCC TEN	90% COBOL 10% Assembler	Database, Shared secontary index, Data set group, Index data fields, Segment, Index data field list, Lchild, Field, Program, Job, PSB application, Module, and 21 more for message formatting and communications

Figure 23--Features of Commercial DD/DS Packages (continued)

SYSTEM NAME	EXPANSIBILITY	STATUS CAPABILITY
DATA CATALOGUE 2	Define additional entities relationships, and attributes. Additional entities used with IMS, TOTAL, DMS, ADABAS, OMS-1100, S2000/80, and IDMS	Status codes: Existing, Proposed, Obsolete, User-defined, and version numbers
DATAMANAGER	Three additional entity structures supported: entity types, attributes, and relationships based on existing ones. Also, additional entity types for each DBMS it supports	Version numbers. Status facility allows partitioning by time, status project etc.
PRIDE/LOGIC	None. All metadata relates to PRIDE methodology	One status for each of nine development phases, modification, improvement, active
LEXICON	Provides capability of creating conventional file and database oriented data descriptions from existing dictionary database entities	Status codes: Proposed, Approved, Concurrent, Effective
EDICT	None	None
TIS DIRECTORY	None	User-defined statuses
ADABAS DATA DICTIONARY	Define additional entities attributes, and relationships, via changes to the D/D schema	No status codes or version numbers. Status capability via separate D/Ds

Figure 23--Features of Commercial DD/DS Packages (continued)

SYSTEM NAME	EXPANSIBILITY	STATUS CAPABILITY
DATA CONTROL SYSTEM (Cincom)	Define additional entities attributes, and relationships, through MANTIS	Test/production versions for the system entities. No status facility
DATA CONTROL SYSTEM (DCS)	Define additional attributes	None
DDS 1100	Define additional entities relationships, and attributes	Proposed, Approved, Obsolete, Active, All, and versions Test, Training, All, Production and User-defined
DATA DICTIONARY (DDS)	None	Operational/User-defined codes and version numbers
DATA DICTIONARY (ADR)	Define additional entities relationships and attributes	Production, History and version numbers
DB/DC DATA DICTIONARY	Define additional entities relationships and attributes	Production, Installed and User-defined
DICTIONARY/204	Define additional entities relationships and attributes	None
Integrated Data (IDD) Dictionary	Define new attributes; full Expansibility planned	Production, and and Historic, and version numbers
XDD DATA DICTIONARY	Define additional entities relationships and attributes	Status codes: Test Production, Load, Development and Obsolete and version numbers
UCC TEN	None	Test and Production status with 256 sides

Figure 23--Features of Commercial DD/DS Packages (continued)

SYSTEM NAME	ON-LINE QUERY	USER-DEFINED REPORTS	SECURITY LEVELS
DATA CATALOGUE 2	YES	Customazation via macro routines; additional reports via call and file extraction capabilities. New reports erquire user software	Three types of password; entity type, and command levels
DATAMANAGER	YES	Customazation via macro routines; additional reports via call and file extraction capabilities. New reports rely on the user. Interface facility	Security assigned at three levels; access, alter and remove
PRIDE/LOGIC	YES	Via extraction facilities	Function, and entity type levels
LEXICON	YES	Via extraction facilities (LEXICON Access Module	Use of D/D special security module;Security assigned at three levels; user supplied password
EDICT	YES	Via the user-defined language	Entity type and others via user defined security routines
TIS DIRECTORY	YES	Via comprehensive retrieval component	Command level
ADABAS DATA DIRECTORY	YES	Via NATURAL , a program development facility	Entity, attribute, attribute value, and function (read and write

Figure 23--Features of Commercial DD/DS Packages (continued)

SYSTEM NAME	ON-LINE QUERY	USER-DEFINED REPORTS	SECURITY LEVEL
DATA CONTROL SYSTEM (Cincom)	YES	Through the Socrates Report Writer	Password facility; security at element level may be specified; additional security through "user exit"
DATA CONTROL SYSTEM (DCS)	YES via QLP	Report options	None
DDS 1100	YES via QLP	Via QLP (Sperry Univac product)	Command and entity occurence
DATA DICTIONARY (DDS)	YES	Report options via the SELECT clause	Entity type, function, user, and operational status
DATA DICTIONARY (ADR)	YES	Through DATAREPORTER	Entire occurence
DB/DC DATA DICTIONARY	YES	Via GIS	Sign on, status, and entity type
DICTIONARY/ 204	YES	Via user-language	Login; Security levels planned
Integrated DATA (IDD) Dictionary	YES	Customazation through changing of parameters; new reports via CULPRIT	User view and record level
XDD DATA DICTIONARY	YES	Via Report Writer	Element entity and command
UCC TEN	YES	Report parameters	Command; more added via security tables

Figure 23--Features of Commercial DD/DS Packages (continued)

SYSTEM NAME	DEPENDENT DBMS OR FILE ORGANIZATION	SOFTWARE INTERFACE PACKAGES
DATA CATALOGUE 2	COBOL relative files	COBOL, PL/I
DATAMANAGER	VSAM or BDAM	COBOL, PL/I, Assembler, ADABAS, IDMS, IMS, DL/I Methodology interface, TOTAL (DDL processor), S2000
PRIDE/LOGIC	COBOL relative files	COBOL, PL/I, JCL, ADF (design aid), IDS II, IMS DL/I
LEXICON	IMS, IDMS, TOTAL	IMS, OS files, IDMS, TOTAL
EDICT	Sequencial File or INQUIRE DBMS	INQUIRE (DDL processor)
TIS DIRECTORY	TIS-DBM	DBM (DBCS), QUERY COMPONENT (Query processor), COMPREHENSIVE RETRIEVAL(Report Writer)
ADABAS DATA DIRECTORY	ADABAS	COBOL,PL/I,NATURAL, ADAWAN (DDL processor), Assembler, ADAMINT (Preprocessor),ADASCRIPT (query processor),ADACOM (report writer)
DATA CONTROL SYSTEM (Cincom)	TOTAL	TOTAL (DDL processor), COBOL
DDS 1100	DMS-1100	DMLP (preprocessor), DMS-1100 (DDL processor)

Figure 23--Features of Commercial DD/DS Packages (continued)

SYSTEM NAME	DEPENDENT DBMS OR FILE ORGANIZATION	SOFTWARE INTERFACE PACKAGES
DATA DICTIONARY (DDS)	IDMS (ICL version)	IDMS (DDL Processor) COBOL, COBOL programs
DATA DICTIONARY (ADR)	DATACOM/DB	DATACOM/DB (DDL Proces.) DATAREPORTER (Report Writer), COBOL, DataQuery (Query Proc.), Datacom/DL (Preproces.>, LIBRARIAN
DB/DC DATA DICTIONARY	IMS or DOS PL/I	COBOL, PL/I, Assembler, MARK IV, IMS (DDL Proc.) DBDA (design aid),GIS (Report Writer), other software
Integrated DATA (IDD) Dictionary	IDMS	IDMS (DDL processor) DML Processor (preproc.) CULPRIT (report writer) OLQ (query processor) IDMS-DC (TP monitor)
XDD DATA DICTIONARY	System 2000/80	System 2000/80 (DDL Processor), COBOL, COBOL PLEX (Preproc.)
UCC TEN	IMS HIDAM Databases	IMS (DDL Processor and System generator), MFS (Terminal Monitor), COBOL, PL/I, Assembler, GIS (General Information System) ADF (Batch Program generator)

Figure 23--Features of Commercial DD/DS Packages (continued)

SYSTEM NAME	INTERACE METHODS/OPTIONS	USERS/COMMENTS
DATA CATALOGUE 2	File extraction/Prefix, Suffix, level number, sequence numbers, and line identifier, include comments and initial values.	Approximately 250 users
DATAMANAGER	File extraction, interface commands/Replace, delete and insert editing options, include comments initial values, and condition names	Approximately 600 users
PRIDE/LOGIC	File extraction, interface commands/	Approximately 300 users
LEXICON	File extraction generated by assembly language, COBOL, or PL/I programs, by means of a transaction/	In January 1980 Andersen, A. and Co. made the decision to with-draw LEXICON from the market place. Support of the system ends 1985
EDICT	File extraction/	Approximately 150 users
TIS DIRECTORY	Interface commands/ Examine and update user views, save the RDL state-ments in the D/D	Approximately 10 users. Integrated pre-compiler planned
ADABAS DATA DIRECTORY	File extraction/Prefixes, alternate functions, sequence numbers	Approximately 1000 users

Figure 23--Features of Commercial DD/DS Packages (continued)

SYSTEM NAME	INTERFACE METHODS/OPTIONS	USERS/COMMENTS
DATA CONTROL SYSTEM (Cincom)	File extraction/More direct interaction if used with other Cincom products	Approximately 130 users. It is also known as DATA DICTIONARY
DDS 1100	Interface commands/	Approximately 30 users
DATA DICTIONARY (DDS)	File extraction, dynamic/ Renaming capability, syntactic check on names, inclusion of comments, initial values and condition names, prefix and suffix capability	Approximately 70 users
DATA DICTIONARY (ADR)	File extraction, interface commands/includes comments	Approximately 100 users
DB/DC DATA DICTIONARY	File extraction, interface commands/Prefix, suffix, level numbers, includes comments. Segment lengths and field start positions recalculated via the RECALCULATE-SEGMENT function	Not available
Integrated Data (IDD) Dictionary	Interface commands/	Approximately 100 users
XDD DATA DICTIONARY	File extraction/Sort the generated data descriptions, select structures within views, renaming	Approximately 40 users
UCC TEN	File extraction/suffix, prefix, and replace editing options, level number controls, include comments, condition names, and initial values	Approximately 300 users

Figure 23--Features of Commercial DD/DS Packages (continued)

SYSTEM NAME	HARDWARE REQUIREMENTS
DATA CATALOGUE 2	IBM 360, 370, 30xx, 43xx Univac 1100, Honeywell 66 series
DATAMANAGER	IBM 360, 370, 30XX, 43XX FACOM M series, Siemens 7000
PRIDE/LOGIC	IBM 360, 370, 30xx, 43xx, Bourroughs Honeywell series 60 & 6000, HP-3000, CDC 6600 DEC 10 & 20, VAX, Univac 1100, ICL 1903, Cyber 175, Prime 750
LEXICON	IBM 360, 370
EDICT	IBM 360, 370, 30xx 43xx
TIS DIRECTORY	IBM 360, 370, 30xx, 43xx
ADABAS	IBM 360, 370, 30xx, 43xx
DATA CONTROL (Cincom)	IBM 360, 370, 30xx, 43xx NCR Century & Criterion
DATA CONTROL (DCS)	Univac 1100
DDS 1100	Univac 1100
DDS (ICL)	ICL 2900
(ADR) DATADICTIONARY	IBM 360, 370, 30xx, 43xx
DB/DC DATA DICTIONARY	IBM 360, 370, 30xx, 43xx
DICTIONARY/204	IBM 360, 370, 30xx, 43xx,
Integrated (IDD)	IBM 360, 370, 30xx, 43xx,
XDD (Intel)	IBM 360, 370, 30xx, 43xx CDC 6000, 70, 170, Univac 1100
UCC TEN	IBM 360, 370, 30xx, 43xx

Figure 23--Features of Commercial DD/DS Packages

B. PROBLEMS AND WEAKNESSES OF DD/DS PACKAGES

Data Dictionary/Directory Systems have not been available for a long time. Nevertheless, organizations began to use them more and more; gaining control and experience from their usage. The commercially available DD/DS packages are not presently capable of providing all functions envisioned for their use. This is not to say that the DD/DSs are worthless, just that there is much room for improvement.

Immediate observations that can be made concerning their capabilities and usage in organizations are the following:

(1) The majority of the commercially available DD/DSs are oriented for use toward a particular DBMS.

(2) There is a broad divergency of opinion as to their scope. On the one hand, the scope of the DD/DS may be quite narrow, covering only the database definitions of a DBMS. On the other hand, its scope may be quite wide, covering all the data resources of an enterprise. The argument for independence of DD/DS and DBMS is not supported by the fact, that the available DD/DS packages provide the range of functions, which can be expected. Some organizations using early DD/DS have centered its use around the directory function, and the DD/DS has become a database definition interface. Only advanced, well managed and large data processing installations have aquired and use a DD/DS as a documentation tool. Further, only few of these DD/DS installations use it as a tool for resolving data conflicts, and constructing clear and unambiguous data definitions.

(3) There is a lack of generally accepted standards, for what constitutes a good data definition.

(4) There is no standard definition of "data element"

(5) It is not quite clear which important characteristics of data should be recorded in the DD/DS.

(6) There is no accepted discipline of conceptual or logical database design.

Among the problems in the use of data of DD/DSs there is a lack of a comprehensive methodology which ties its use into the SDLC [Ref. 33]. The SDLC approach is considered as a management tool to plan and control systems development efforts. DD/DSs are considered as technical support tools associated with database management systems. On the one hand the DD/DS has to support and measure the progress of a project's development life cycle. On the other hand, the SDLC must specify activities in terms of the DD/DS. Evidences of this situation are considered in the following:

- No standard definitions of entity types.
- Methodologies used, such as, structured analysis, refer primarily to the flow of data rather than the flow of control and usage of a DD/DS.
- Most DD/DS are oriented toward recording data (definitions and structures), as they are in a physical database or COBOL file. They are not oriented toward the highly dynamic design process itself.
- A few commercially available DD/DSs, provide graphic representations of the relationships between the object identifiers, which is very important for logical design methodologies.
- Only a few of the current DD/DSs are integrated, and most of them act like passive DD/DSs.

The above mentioned problems do not imply that the DD/DS is not of great promise in assisting the management of the information resource. It will be undergoing major change during the years to come. It will be more integrated, its use will be an accepted part of the systems development and maintenance life cycle. Its use among non-DP staff will be much more extensive than at the present time. To some degree today, because of product deficiencies and the lack of well established methodologies for usage, DD/DS seem to fail in practice to achieve its goals.

C. EXPECTATIONS ON THE FUTURE OF DD/DS

The evolution of DD/DS since 1970s, like that of all computer software, was the result of two driving forces: The need of the user, and the foresight of the developer. Each of them has an active role to play and each has its limitations.

The need of the user today and in the future can be characterized by broader demand for easy-to-use facilities, greater need for distributed processing functions, cheaper systems, and much more widespread use of database technology. Developers on the other hand, tend to see the broader view of a system, but occasionally they overlook important aspects. This is due to a lack of exposure to day-to-day business situations. Nevertheless, a current trend in the computer industry is the availability of cheaper, smaller, and more effective systems. The usage of DD/DS products in the future will be a balance of these forces. It will be determined by the user demand for facilities, and by user acceptance of the facilities fabricated by developers.

Projections on the future of DD/DS can be classified into two categories: Short-term developments, that are specific and concrete, and long-term that are often transparent. Short-term projections are:

- More powerful query and analysis capabilities.
- More user-friendly interfaces.
- Greater integration of DD/DS into actual life cycle management.
- New structural techniques for modeling purposes. This will result in the ability of the DD/DS to deal with richer semantic constructs.

However, the major topics for consideration is in the long-term development of DD/DS, and its usage in:

- Integration and evaluation into the IRM concept.
- Mini-and microcomputer applications; and
- Distributed systems (networks), consisting of several different types of DBMSs, file managers, text editors that run on computers from different manufactureres.

Into this environment, the Data Dictionary/Directory System would act as a driver of the entire system.

V. EVALUATION/SELECTION CRITERIA FOR DD/DS ACQUISITION

When management decides to implement a DD/DS the first consideration is whether to purchase a commercially available software package or develop one in-house. This make versus buy decision is based on a combination of many factors: hardware, software, user acceptance, planning for data administration, cost, and many others. In summary, there are technical and economic factors, as well as, organization needs. Nevertheless, selection of a DD/DS must be based primarily on the needs of the organization.

To build an in-house design and implement a DD/DS requires three main resources: money, technical expertise, and time. If an organization has all three, then it would be possible to build a DD/DS. Again, there is some evidence that inhibits the purchase of commercial DD/DS packages: Organizations that have non-IBM equipment have a minimum number of DD/DSs to choose from. Organizations with minicomputers and microcomputers have an even more limited selection. Organizations (scientific) that need a very flexible set of entity types and attributes are not supported by the commercially available DD/DSs. If no commercial DD/DS exists that will run with existing software (e.g., operating system or compiler), then there may be no other option than to build a DD/DS. An organization considering to build its own DD/DS can gain in reducing subsequent operational and maintenance costs; and can better match user needs than a generalized commercial system.

Even though, there are some important constraint factors that suggest the alternative way of buying one:

- The design and implementation of a DD/DS is a non-trivial task. It needs a great deal of effort, time, and costs a large amount of money, especially if the system must be continually updated.

- Today, DD/DS is considered as a highly sophisticated product, providing primarily high reliability. In an in house developed DD/DS, it is possible that undetected errors might be resident for longer periods of time than in a commercially available system.

- To keep track with the technology, DD/DS needs continuous enhancements, that will improve and increase its effectiveness in an organization. Such enhancements are very difficult to be done for an in-house system, and are better provided by outside software products.

These reasons, and also the commercial availability of an increasing number of DD/DS packages suggest that the buy alternative should be considered as the most preferable.

Before initiating the selection process, an organization must determine if a DD/DS is justified, based on an economic analysis of costs and benefits or savings of implementing the system. As stated earlier, the DD/DS is not an integral part of a database management system in most of the packages available today. As a result, the decision to buy a DD/DS package should to be made independently of the DBMS. It must be mentioned here that if an organization also needs a DBMS, the ideal time to aquire a DD/DS package, is before selection of a DBMS. This will allow the organization to set in motion a number of administrative and conceptual adjustments that will assist and pave the way for database.

As in any economic analysis, determining an actual dollar value for savings or benefits may be extremely difficult and is a subjective judgement. Benefits may be expressed in terms of savings, cost avoidance, improved performance, and hidden opportunities. Fig. 24 lists some aspects of costs and savings/benefits which should be considered in the economic analysis. Fig. 24 is not an all inclusive list; management should determine actual cost/benefit categories to be considered applicable to an organization.

COSTS	BENEFITS/SAVINGS
Acquisition	System Design and Development
Data Administration Staff	System Maintenance
Hardware costs	Data Redundancy
Start-up costs	Database Creation
Data collection costs	Auditing Information Resource
Maintenance	Improved Communications
Application System Changes	
User Training	

Figure 24--Costs and Savings/Benefits of DD/DS

Selection of a DD/DS should be based on who will use the system and how it will be used, rather than what is the most technologically innovative system in the market. Lefkovits, et al., [Ref. 34], recommend the following selection and evaluation process:

- Determine requirements; classify which requirements are mandatory, and which are the desirable features, with a corresponding point scale indicating importance.

- Develop a list of features to be used in the evaluation of DD/DS.

- Determine a mapping from requirements to features; multiple mappings may be possible.

- Compare features provided by commercially available systems to each mapping to determine if a system qualifies for further consideration.

- Compare those systems which qualify for the degree of compliance of any available desirable features, assigning a point value.

- Sum point values assigned to desirable features of qualified systems to select the DD/DS which best meets the requirements.

We cannot assure that this process does not include some risk, since subjective judgement on the part of management is involved. The wrong system may still be selected for many reasons: improper determination of requirements; usually due to a lack of user involvement in the selection process; improper qualification of features, due to technical bias of selection team; inconsistent evaluation of the system, due to different members of the selection team, as well as, a lack of a well-defined measurement methodology; and undue emphasis on features needed in the future, but not at the time of implementation, which could result in user dissatisfaction with an unnecessarily complex system.

The system being selected and implemented, should be evaluated periodically to determine its performance. Often the requirements and needs of an organization change, requiring a reevaluation of the DD/DS to determine the new and/or changed demands. If the DD/DS no longer meets these new requirements in an acceptable fashion, a new system must be evaluated and selected.

VI. CONCLUSION

The explosive proliferation of computer usage in the last ten years has led organizations to increase their Data Processing Activities. As more and more activities, vital to an organization, become automated, the actual processing of data and information becomes more and more strategic to that organization. Today, corporate management is becoming aware of an important asset which, until recently, has been virtually ignored. The asset is data. The idea of data being a corporate asset is relatively new, and has developed along with the influence and proliferation of computers in organizations. Data must be stored, managed, protected, and and retrieved effectively and efficiently, as does any other critical organizational resource. To accomplish these tasks, several kinds of management software tools have been developed. One tool which management can utilize is the DD/DS.

DD/DS is presently in an evolutionary state. Originally started as a tool for description and documentation within a database, it was soon recognized that the productivity services, that the DD/DS could provide, were part and parcel of data resource control. However, DD/DS implementations have lagged somewhat behind that of earlier developed DBMS. The confusion regarding scope, definition and integration of currently available DD/DS somewhat hinders the effective and widespread implementation of DD/DS. Some of the functions which DD/DS purport to possess are still theoretical in nature. Also, a lack of user education leads to erroneous or misdirected use of DD/DS. Nevertheless, the increasing interest and attention in this area has led to improvements in DD/DS. Recently, the data management community has begun to recognize the DD/DS as a more general tool for data resource management. It pertains not only to database systems, but increasingly, to non-database systems as well, and it is further broadened by its support for job streams, structured systems, on-line environments, and systems design.

One area of computer processing and data resource management where the DD/DS has been found to be a useful tool is in the SDLC. It can be used to support various activities throughout the system development process including the maintenance and operation phase. Designers, analysts, and persons who are involved in system development and operation should use and rely on the DD/DS. Recent experience has shown that information systems developed with the aid of a DD/DS tend to exhibit fewer errors that need correcting, as well as errors which are not as serious, compared to systems developed in such an environment. This result, in consideration of the resources that are often required for system maintenance make the DD/DS a useful tool in the System Development Life Cycle.

LIST OF REFERENCES

1. Senn, J. A., Information Systems in Management, 2d ed., p. 18, Wadsworth Publishing Company, 1978.

2. ibid., p. 20.

3. Plagman, B. K., Data Dictionary / Directory System : A Tool for Data Administration and Control, p. 42, in Auerbach Database Management Series, Portfolio 22-01-02, 1977.

4. O'Brien, J. A., Computers in Business Management : An Introduction, 3d ed., p. 23, Richard D. Irwin, Inc., 1982

5. Van Duyn, J. A., Developing a Data Dictionary System, pp. 5-6, Prentice-Hall, Englewood Cliffs, New Jersey, 1981.

6. Kroenke, D. M., Database Processing, 2d ed., p. 11, Science Research Associates, Inc., 1983.

7. ibid., p. 178.

8. U.S. Department of Commerce, National Bureau of Standards, Special Publication 500-92, Database Directions: Information Resource Management--Strategy and Tools, Goldfine, A. H., ed., p. 50, 1982.

9. U.S. Department of Commerce, National Bureau of Standards, Special Publication 500-28, Database Administration: Concepts, Tools, Experiences and Problems, by Leong-Hong, B. and Marron, B., p. 4, March 1978.

10. Atre, S., Database: Structured Techniques for Design, Performance, and Management, p. 54, Wiley, New York 1980

11. Kroenke, p. 193.

12. Rullo, T. A., Advances in Database Management, Vol. 1, pp. 39-40, Heyden & Son, Inc., 1980.

13. Ross, R. G., Data Dictionaries and Data Administration, pp. 376-380, Amacom, 1981.

14. U.S Department of Commerce, National Bureau of Standards, Special Puplication 500-3, Technical Profile of Seven Data Element Dictionary/Directory Systems, by Leong-Hong, B., and Marron, B., pp. 1-8, February 1977.

15. National Bureau of Standards, Special Publication 500-28, pp. 127-128.

16. Cardenas, A. F., Database Management Systems, p. 462, 1979.

17. Leong-Hong, B. W., and Plagman, B. K., Data Dictionary/Directory Systems: Administration, Implementation and Usage, John Wiley & Sons, Inc., p. 16., 1982.

18. Van Duyn, pp. 22-23.

19. Allen, F. W., Loomis, M. E. S., and Mannino, M. V., The Integrated Dictionary/Directory System", ACM Computing Surveys, v. 14, n. 2, p. 247, June 1982.

20. Leong-Hong, and Plagman, p. 10.

21. ibid., pp. 20-22.

22. Allen, et. al., p. 250.

23. Leong-Hong, and Plagman, pp. 78-87.

24. Allen, et. al., pp. 249-250.

25. Perry, W. E., Data Base Management: Control in a Data Base Environment, Q.E.D Information Sciences, Inc., Monograph Series, No 10, pp. 43-45, 1980.

26. Allen, et al., p. 251.

27. The British Computer Society Data Dictionary Systems Working Party Report, Data Base, v. 9, n. 2, pp. 3-4, Fall 1977.

28. Leong-Hong, and Plagman, p. 29.

29. Lefkovits, H. C., Sibley, E. H., and Lefkovits, S. L., Information Resource/Data Dictionary Systems, Q.E.D Information Sciences, Inc., pp. 1.18-1.20, 1983.

30. ibid., pp. 3.1-A.97.

31. Allen, et al., pp. 276-286.

32. Van Duyn, pp. 156-165.

33. Curtice, R. M., Data Dictionaries: An Assessment of Current Practices and Problems, Proceedings of the

Seventh International Conference on Very Large Data Bases, p. 565, Arthur D. Little, Inc., 1981.

34. Lefkovits, et al., pp. 1.38-1.41.

BIBLIOGRAPHY

Curtice, R. M., and Dieckmann, E. M., "A Survey of Data Dictionaries," Datamation, v. 27, n. 3, March 1981.

Elbra, R. A., and Hardy, N., "Cost Justification of Data Dictionaries," Computer Bulletin, ser. 2, n. 25, September 1980.

Kennedy, R. M., "Data Dictionary Experience at West Midlands Gas," Database Journal, v. 11, n. 1, 1981.

Kreitzer, L. W., Data Dictionaries--the Heart of IRM Infosystems, v. 28, n. 2, February 1981.

Morehen, H. L., Data Dictionaries, Evaluation and Implementation, Australian Computer Bulletin, v. 6, n. 10, November 1982.

Patel, H., The Data Dictionary as a Tool for Systems Auditability and Control, Computer Audit and Control-State of the Art Report, Infotech, 1980.

Schussel, G., "The Role of Data Dictionaries," Datamation, v. 23, n. 6, June 1977.

U.S. Department of Commerce, Federal Information Processing Standard Publication 76, Guideline for Planning and Using a Data Dictionary System, 20 August 1980.

U. S. Department of Commerce, National Bureau of Standards Special Publication 500-16, Survey of Eleven Government Developed Data Element Dictionary/Directory Systems, 1977.

Windson, A. T., ed., Using the ICL Data Dictionary. Proceedings of the User Group, Shiva Publishing, 1980.

INITIAL DISTRIBUTION LIST

		No. Copies
1.	Defence Technical Information Center Cameron Station Alexandria, Virginia 22304-6145	2
2.	Library, Code 0142 Naval Postgraduate School Monterey, California 93943-5100	2
3.	Computer Technology Curricular Office Naval Postgraduate School Code 37 Monterey, California 93943-5100	1
4.	Chairman, Code 54 Gk Department of Administrative Sciences Naval Postgraduate Scool Monterey, California 93943-5100	1
5.	Associate Professor N. R. Lyons, Code 54Lb Department of Administrative Sciences Naval Postgraduate School Monterey, California 93943-5100	1
6.	LCDR Barry Frew, Code 54 Fw Department of Administrative Sciences Naval Postgraduate School Monterey, California 93943-5100	1
7.	EMBASSY OF GREECE Naval Attache 2228 Massachusetts Ave., N. W. Washington, D. C. 20008	6

600- 197

Thes
D72
c.1

216319

Thesis
D7283 Drakoulis
c.1 Data Dictionary/Directory Systems, a tool in Systems Development Life Cycle.

4 MAR 87
6 AUG 90

216319

Thesis
D7283 Drakoulis
c.1 Data Dictionary/Directory Systems, a tool in Systems Development Life Cycle.

thesD7283
Data Dictionary/Directory Systems, a too
3 2768 000 64807 5
DUDLEY KNOX LIBRARY